POPLAR OXIDOREDUCTASES INVOLVED IN THE OXIDATIVE STRESS RESPONSE: A CRYSTALLOGRAPHIC SNAPSHOT TOWARDS THE UNDERSTANDING OF THE CATALYTIC MECHANISMS

by

KOH CHA SAN

Thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

May 2008

ACKNOWLEDGEMENTS

First, I am very grateful to the commission board members of my thesis who accept and willing to read my thesis. I would like to thank two of my main supervisors, Prof. Catherine Corbier from France, and Associate Prof. Dr Razip Samian from Malaysia, who spent a lot of time reading my thesis and gave me a lot of useful suggestions and criticisms, and also for being patient with me through all the times. I really appreciate the guidance and assistances that they have provided and very grateful them to accept me as their student under the UHP-USM cotutelle program. I also like to express my gratitude to Dr Claude Didierjean who shared with me a lot of his expertise and research insight. He quickly became for me the role model of a successful researcher in the field. I could not have imagined having a better mentor, and without his enthusiasm, his inspiration, his unselfish passion and his encouragement, I would never have finished.

The members of my supervisory committee provided with many useful suggestions and comments to me. This includes Prof. Omar Shawkataly, Associate Prof. Dr. Sheila Nathan and Dr. André Aubry. I want to thank Dr. Fréderique Favier and Dr. Guillermo Mulliert, who taught me the knowledge of protein X-ray crystallography and protein modeling, which they have provided me a much broader understanding of these subjects. Many thanks go to the director of LCM^3B Prof. Claude Lecomte who welcomed me in the laboratory and to the present and past LCM^3B members particularly Dr. Moutsé Ranaivoson, Sébastien Moniot, Dr. Hélène Dubourg, Dr. Fabien Pascale, Dr. Sébastien Lebègue, Sandrine Mathiot, Alexandre Bouché, Dorothee Liebschner, Mikael Elias etc. who offered countless aides, discussions and provided a friendly working environment in the laboratory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCD	Charge-Coupled Device
COX	Cytochrome c oxidase
DHAR	Dehydroascorbate reductase
DTT	Dithiothreitol
FT	Fourier Transform
FTR	Ferredoxin/Thioredoxin Reductase
Gpx	Glutathione peroxidase
GR	GSH-reductase
Grx	Glutaredoxin
GSH	Glutathione
GST	Glutathione-*S*-transferase
HED	β-hydroxyethyl disulfide
IR	Infrared
MAD	Multi-wavelength anomalous diffraction
MAPK	Map Kinase
MAPKKK	Map Kinase Kinase Kinase
MD	Molecular Dynamics
MIR	Multiple Isomorphous Replacement
MPD	2-Methyl-2,4-Pentanediol
NMR	Nuclear Magnetic Resonance
NTR	NADPH/ Thioredoxin reductase
OB fold	Oligonucleotide-binding fold

'DB	Protein Data Bank
'DI	Protein Disulfide Isomerase
'EG	Polyethylene glycol
'ICOT	Protein Kinase C-interacting cousin of thioredoxin homology domain
'rx	Peroxiredoxin
ROS	Reactive Oxygen Species
SAD	Single-wavelength anomalous diffraction
SIRAS	Single isomorphous replacement with anomalous scattering
S*n*B	Shake-and-Bake
TLS	Translation, Libration, Screw
Trx	Thioredoxin
TrxR	Thioredoxin reductase
VMD	Visual Molecular Dynamics
WT	Wild-type

OKSIDOREDUKTASE DARIPADA KAYU KERAS JENIS POPULUS (POPLAR) YANG TERLIBAT DALAM RESPONS TERHADAP TEKANAN OKSIDATIF: KAJIAN KRISTALOGRAFIK KE ARAH PEMAHAMAN MEKANISME PEMANGKINAN ENZIM

ABSTRAK

Penghasilan species oksigen reaktif (ROS) merupakan akibat daripada persekitaran sentiasa dalam keaadaan aerobik yang tidak boleh dielakkan dan ia boleh disingkirkan oleh beberapa jenis sistem enzim sebaik sahaja dihasilkan. Tiga jenis oksidoreduktase (Glutathion peroksidase, Gpx; thioredoksin, Trx; and glutaredoksin, Grx) daripada *Populus trichocarpa* × *deltoides* (kayu keras jenis poplar) telah dicirikan melalui pendekatan biokimia dan kristalografi X-ray. Gpxs terdiri daripada sekumpulan enzim yang mengawal tahap ROS dalam sel dan melindunginya daripada kerosakan oksidatif. Dalam kajian ini, saya telah mencirikan struktur kristal bagi Gpx5 poplar (PtGpx5) dalam bentuk terturun dan teroksida. Perbandingan antara kedua-dua struktur redoks menunjukkan bahawa perubahan konformasi yang drastik adalah satu keperluan untuk menghampirkan dua residu sistin yang berjauhan supaya membentuk ikatan disulfida intra-molekul. Trxs ialah sekumpulan enzim yang mengawalatur pelbagai protein melalui proses penurunan thiol-disulfida. Objektif kajian ini adalah untuk menerangkan proses pemangkinan sejenis isoform Trx baru poplar, PtTrx*h*4, memandangkan enzim ini hanya boleh diturunkan oleh Grx dan bukannya oleh sistem NADPH:thioredoksin reduktase yang biasanya digunakan oleh Trx. PtTrx*h*4 mengandungi tiga sistin; satu di bahagian terminal N protein (Cys4) dan dua lagi (Cys58 and Cys61) di tapak aktif tipikal Trx (WC_1GPC_2). Analisis dua struktur kristal PtTrx*h*4 dalam kajian ini, jenis liar dan mutan C61S, membolehkan kita mencadangkan satu mekanisme pemangkinan empat langkah yang juga disokong oleh kajian enzimatik. Grxs pula adalah protein redoks yang menggunakan elektron daripada GSH untuk pemangkinan proses penukaran thiol-disulfida. Di sini, saya mencirikan struktur PtGrxS12 yang berkompleks dengan GSH, di mana ia adalah struktur pertama dari kumpulan Grx tumbuhan jenis sub-kelas 1 yang memiliki tapak aktif atipikal $_{28}WCSYS_{32}$. Protein ini mempunyai sistin tambahan (Cys87) dan peranan residu tersebut masih tidak diketahui. Maklumat tentang tapak ikatan GSH juga dirumuskan dalam kajian ini. Semua struktur protein yang dicirikan di sini mendalami pengetahuan kita tentang pemangkinan redoks dalam sel umbuhan dan interaksi enzim-substrat.

POPLAR OXIDOREDUCTASES INVOLVED IN THE OXIDATIVE STRESS RESPONSE: A CRYSTALLOGRAPHIC SNAPSHOT TOWARDS THE UNDERSTANDING OF THE CATALYTIC MECHANISMS

ABSTRACT

The production of reactive oxygen species (ROS) is an unavoidable consequence of living in an aerobic environment and once produced, it can be removed by several different enzyme systems. Three oxidoreductases (glutathione peroxidase, Gpx; thioredoxin, Trx and glutaredoxin, Grx) from *Populus trichocarpa* × *deltoides* (poplar tree) were characterized using biochemistry and X-ray crystallography approaches. Gpxs are a group of enzymes that regulate the levels of ROS in cells, and protect them against oxidative damage. In this study, I have determined the crystal structures of the reduced and oxidized form of poplar Gpx5 (PtGpx5). Comparison of both redox structures indicates that a drastic conformational change is necessary to bring the two distant cysteine residues together to form an intramolecular disulfide bond. Trxs are a family of ubiquitous enzymes which regulate various protein partners through the thiol-disulfide(s) reduction. The aim of this study is thus to precisely describe the catalytic mechanism of a new isoform of Trx that has been characterized in poplar, PtTrx*h*4, since it has been demonstrated recently to be reduced by Grx but not by the typical NADPH:thioredoxin reductase reducing system. PtTrx*h*4 contains three cysteines; one localized in an N-terminal extension (Cys4) and two (Cys58 and Cys61) in the usual Trx active site (WC_1GPC_2). Analyses of two PtTrx*h*4 crystal structures solved in this study, wild-type and C61S mutant, allow us to propose a four-step disulfide cascade catalytic mechanism in accordance with enzymatic studies. Grxs are highly conserved redox-proteins that utilize electrons from GSH particularly to catalyze thiol-disulfide exchange reactions. Here, I present the structure of glutathionylated PtGrxS12, the first structure of plant Grx of subclass 1 with an atypical $_{28}WCSYS_{32}$ active site. This protein possesses an additional cysteine (Cys87) in which the role of this extra active-site cysteine remains obscure. Details of the GSH binding site are summarized in this study. Protein structures solved here shed lights to our understanding of the redox mechanism in plants and to the enzyme-substrate interactions.

ÉTUDE CRISTALLOGRAPHIQUE DES OXYDORÉDUCTASES IMPLIQUÉES DANS LA RÉPONSE AU STRESS OXYDATIF CHEZ LE PEUPLIER POUR UNE MEILLEURE COMPRÉHENSION DES MÉCANISMES CATALYTIQUES

RÉSUMÉ

La production des espèces réactives de l'oxygène (ROS) est une conséquence inévitable de la vie dans un milieu aérobie. Une fois produites, ces espèces peuvent être éliminées par différents systèmes enzymatiques. Les structure de trois oxydoréductases (la glutathion peroxydase (Gpx), la thiorédoxine (Trx) et la glutarédoxine (Grx)) de *Populus trichocarpa* × *deltoides* (le peuplier) ont été résolues par diffraction des rayons X. Les Gpxs forment un groupe d'enzymes qui régulent la concentration des ROS dans les cellules, et qui les protègent des effets d'un stress oxydant. Contrairement à leurs homologues d'origine animale, les Gpxs végétales ne dépendent pas du glutathion (GSH) mais des Trx pour leur fonctionnement. De plus, elles possèdent une cystéine dans leur site actif au lieu d'une sélénocystéine. Dans cette étude, j'ai résolu les structures cristallines des formes réduite et oxydée de la Gpx5 de peuplier (PtGpx5) et montré que des changements conformationnels drastiques sont nécessaires pour permettre la formation d'un pont disulfure intramoléculaire. Les Trxs constituent une famille de petites protéines ubiquitaires, régulant diverses protéines cibles par la réduction de leur pont disulfure. Par cette étude structurale, mon objectif était de comprendre précisément le mécanisme catalytique d'une nouvelle isoforme caractérisée chez le peuplier, la PtTrx*h*4, dont la capacité à accepter des électrons de la Grx a été récemment démontrée. A la différence des autres Trxs *h* caractérisées jusqu'ici, la PtTrx*h*4 ne peut pas être réduite par la voie typique qui implique la NADPH-Trx réductase. Cette PtTrx*h*4 contient trois cystéines, la première localisée dans une extension en position N-terminale (Cys4) et deux (Cys58 et Cys61) situées dans le site actif classique (WC_1GPC_2) de la Trx. Les résolutions des structures de l'enzyme sauvage et du mutant C4S sous forme oxydée m'ont permis de proposer un mécanisme catalytique en quatre étapes en accord avec les études enzymatiques. Les Grxs sont des protéines qui utilisent des électrons de GSH en particulier pour catalyser des réactions d'échange de thiol-disulfure. Ici, je présente la structure de la PtGrxS12 (en complexe avec le GSH), la première structure de la Grx végétale sous-classe 1 avec un motif du site actif atypique $_{28}WCSYS_{32}$. Cette protéine possède une autre cystéine (Cys87), dont le rôle reste obscur. Le site de fixation du GSH est détaillé ultérieurement dans ce manuscrit. Ces structures éclaircissent notre compréhension du mécanisme d'oxydoréduction chez les végétaux et nous permettent de mieux comprendre les interactions enzyme-substrat.

Introduction

CHAPTER 1

INTRODUCTION

1.1 An overview of X-ray crystallography and protein crystallography

X-ray crystallography is the science of determining the arrangement of atoms within a crystal from the manner in which a beam of X-rays is scattered from the electrons within the crystal. The method produces a three-dimensional picture of the electron density within the crystal, from which the mean atomic positions, their chemical bond, their disorders and sundry other information can be derived. By definition, a crystal is a solid in which a particular arrangement of atoms (its unit cell) is repeated indefinitely along three principal directions known as the basis (or lattice) vectors, which are not necessarily perpendicular. A wide variety of materials can form crystals; such as salts, metals, minerals, semiconductors, as well as various inorganic, organic and biological molecules, which has made X-ray crystallography fundamental to many scientific fields.

The oldest and most precise method of X-ray crystallography is single-crystal X-ray diffraction, in which a beam of X-rays is reflected from evenly spaced planes of a single crystal, producing a diffraction pattern of spots called reflections. Each reflection corresponds to one set of evenly spaced planes within the crystal. The density of electrons within the crystal is determined from the position and intensity of the various reflections observed as the crystal is gradually rotated in the X-ray beam; this density allows the atomic positions to be inferred. For single crystals of sufficient purity and regularity, X-ray diffraction data can determine the mean chemical bond lengths and angles to within a few thousandths of an Angstrom and to

within a few tenths of a degree, respectively. The data also allow the static and dynamic disorder in the atomic positions to be estimated, which is usually less than a few tenths of an Angstrom. At its most basic level, X-ray crystallography is useful in identifying known materials, characterizing new materials and in discerning materials that appear similar by other experiments. However, X-ray crystal structures have many other applications; for example, they can account for unusual electronic or elastic properties of a material, shed light on chemical interactions and processes, or serve as the basis for understanding enzymatic mechanisms and designing inhibitors of therapeutic roles with hope of preventing diseases. The term "X-ray crystallography" is also sometimes applied to methods that involve X-ray diffraction from polycrystalline materials, such as powders of small crystals studied by X-ray powder diffraction.

Early protein crystallographers, proceeding by analogy with studies of other crystalline substances, examined dried protein crystals and obtained no diffraction patterns. Thus X-ray diffraction did not appear to be a promising tool for analyzing proteins. In 1934, J. D. Bernal and Dorothy Crowfoot (later Hodgkin) measured diffraction from pepsin crystals still in the mother liquor. Bernal and Crowfoot recorded sharp diffraction patterns, with reflections out to distances in reciprocal space that inversely correspond in real space to the distances between atoms. The announcement of their success was the birth announcement of protein crystallography. Ever since, X-ray crystallography has been the most prolific technique for the structural analysis of proteins and protein complexes, and is still the "gold standard" in terms of accuracy and resolution (Russell et al. 2004). Each new structure provided a specific step forward, each a harder puzzle due to its size, the resolution needed or any number of other complicating factors. To list but a few,

GroEL-GroES (Xu et al. 1997), F_1-ATPase (Abrahams et al. 1994, Leslie & Walker 2000), the nucleosome (Davey et al. 2002, Schalch et al. 2005), a bacterial potassium ion channel (Doyle et al. 1998), photosystem I (Ben-Shem et al. 2003, Amunts et al. 2007), the ribosome (Ban et al. 2000), the plasma membrane proton pump (Pedersen et al. 2007) and a bacterial multidrug ABC transporter (Dawson & Locher 2006), certainly make the milestones seem closer as the protein structures proliferate. Nevertheless, the number of structures of macromolecular assemblies solved by X-ray crystallography is still quite small compared to that of the individual proteins and it will probably be many years before we have a complete repertoire of high-resolution structures for the hundreds of complexes in a typical cell. This discrepancy is due mainly to the difficult production of sufficient quantities of the sample and its crystallization (Russell et al. 2004).

1.1.1 Nobel prize winners associated with crystallography

The beauty and regularity of crystals impressed people such an extent that, in the past, crystals were regarded as products of nature with mysterious properties. Scientific investigation of crystals started in 1669, when Nicolaus Steno, a Dane working as a court physician in Tuscan, proposed that during crystal growth, the angles between the faces remained constant. For a given crystal form, individual crystals might differ in shape (for example, in the development of their faces), but they always have identical angles between the same faces.

Before the famous first X-ray crystallographic diffraction experiment by von Laue, Friedrich and Knipping in 1912, the internal regularity of a crystal was suggested but never proven. X-ray crystallography has dramatically changed this

situation. The idea that crystals could be used as a diffraction grating for X-rays arose in 1912 in a conversation between Paul Peter Ewald and Max von Laue in Munich. Ewald had proposed a resonator model of crystals for his thesis, but this model could not be validated using visible light, since the wavelength was much larger than the spacing between the resonators. Von Laue realized that electromagnetic radiation of a shorter wavelength was needed to observe such small spacings, and suggested that X-rays might have a wavelength comparable to the unit-cell spacing in crystals. Working with Friedrich and Knipping, they shined a beam of X-rays through a sphalerite crystal and recorded its diffraction (a large number of well-defined spots arranged in a pattern of intersecting circles around the spot produced by the central beam) on a photographic plate. Von Laue developed a law that connects the scattering angles and the size and orientation of the unit-cell spacings in the crystal, for which he was awarded the Nobel Prize in Physics in 1914. Table 1.1 summarized the Nobel laureates with their contributions in the domain of crystallography.

Crystal structures of proteins began to be solved in the late 1950's, beginning with the structure of sperm whale myoglobin (Kendrew et al. 1960) and hemoglobin (Perutz et al. 1960) by Max Perutz and Sir John Cowdery Kendrew, for which they were awarded the Nobel Prize in Chemistry in 1962. Since then, the progress of protein crystallography has followed an evolutionary rather than a revolutionary path. Many important advances have been achieved on the way, but most of the methods used by Perutz are still valid today and still useful, albeit often in modified versions (Dauter 2005). Since that success, over 50000 X-ray crystal structures of proteins, nucleic acids and other biological molecules have been determined (according to PDB statistics in June 2008).

Table 1.1 Nobel Prize laureates in scientific fields associated with crystallography.

Year	Domain	Nobel Laureate(s)	Findings / Scientific contributions
1901	Physics	W.C.Röntgen	Discovery of X-rays
1914	Physics	M Von Laue	Diffraction of X-rays by crystals
1915	Physics	W.H.Bragg & W.L.Bragg	Use of X-rays to determine crystal structure
1929	Physics	L-V de Broglie	The wave nature of the electron
1937	Physics	C.J.Davisson & G.Thompson	Diffraction of electrons by crystals
1946	Chemistry	J.B.Sumner	Discovery of enzymes that can be crystallized
1954	Chemistry	L.C.Pauling	Research of the nature of the chemical bond and its application to the elucidation of the structure of complex substances
1962	Physiology or Medicine	F.Crick, J.Watson & M.Wilkins	Discovery of the helical structure of DNA Books about the double helix
1962	Chemistry	J.C.Kendrew & M.Perutz	Studies of the structures of globular proteins
1964	Chemistry	D.Hodgkin	Determination of many biochemical substances structures including Vitamin B12
1976	Chemistry	W.N.Lipscomb	Elucidation of the structure of boranes
1982	Chemistry	A.Klug	Development of crystallographic electron microscopy and discovery of the structure of biologically important nucleic acid-protein complexes
1982	Physics	K.G.Wilson	Conceptualizing the theory of critical phenomena in connection with phase transitions
1985	Chemistry	H.Hauptmanm & J.Karle	Development of direct methods for the determination of crystal structures
1988	Chemistry	J.Deisenhofer, R.Huber & H.Michel	Determination of the three-dimensional structure of a photosynthetic reaction centre
1991	Physics	P-G de Gennes	Development of the methods of discovering order in simple systems that can be applied to polymers and liquid crystals
1992	Physics	G.Charpak	Discovery of the multi wire proportional chamber
1994	Physics	C.Shull & N.Brockhouse	Neutron diffraction
1996	Chemistry	R.Curl, H.Kroto & R.Smalley	Discovery of the fullerene form of carbon
1997	Chemistry	P.D.Boyer, J.E.Walker & J.C.Skou	Elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP) and discovery of an ion-transporting enzyme
2002	Chemistry	J.B.Fenn, K.Tanaka & K.Wüthrich	Development of methods for identification and structure analyses of biological macromolecules
2002	Physics	R.Davis Jr., M.Koshiba & R.Giacconi	Pioneering contributions to astrophysics (detection of cosmic neutrinos and the discovery of cosmic X-ray sources)
2003	Chemistry	P.Agre & R.Mackinnon	Discoveries concerning (water and ion) channels in cell membranes
2006	Chemistry	R.D.Kornberg	Studies of the molecular basis of eukaryotic transcription

1.1.2 Integration and applications of crystallography in structural biology today

The development and application of technologies in structural biology (availability of complete genome sequences, automation of cDNA cloning, automated protein expression screens, affinity tags and parallel purification strategies, high-throughput crystallization procedures, usage of third-generation synchrotron, automated sample changing and the use of high-energy sources and CCD detector technology) do accelerate the transition from gene to structure (Dry et al. 2000). X-ray crystallography has benefited from several technological advances in recent years that make the genome-wide protein structure determination a practical reality. Indeed, it is now used routinely by scientists to determine how a pharmaceutical interacts with its protein target and what changes might be advisable to improve it (Scapin 2006). The number of protein structures deposited in the Protein Data Bank now (June 2008) is almost 48 000, with the vast majority (85 %) determined using X-ray crystallographic methods. Among the available protein crystal structures, only 34 % (~14000 structures) of them actually represent unique proteins (structures with similar sequences at 90 % identity were excluded). Thousands of studies describing such structures have been published in the scientific literature, and 14 Nobel prizes in chemistry or medicine have been awarded to protein crystallographers (see Table 1.1).

There is a wide spectrum of experimental and computational methods for the identification and structural characterization of macromolecular complexes. These methods need to be combined into hybrid approaches to achieve greater accuracy, coverage, resolution and efficiency than any of the individual methods. New methods must be capable of generating possible alternative models consistent with

POPLAR OXIDOREDUCTASES INVOLVED IN THE OXIDATIVE STRESS RESPONSE: A CRYSTALLOGRAPHIC SNAPSHOT TOWARDS THE UNDERSTANDING OF THE CATALYTIC MECHANISMS

KOH CHA SAN

UNIVERSITI SAINS MALAYSIA
2008

POPLAR OXIDOREDUCTASES INVOLVED IN THE OXIDATIVE STRESS RESPONSE: A CRYSTALLOGRAPHIC SNAPSHOT TOWARDS THE UNDERSTANDING OF THE CATALYTIC MECHANISMS

by

KOH CHA SAN

Thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

May 2008

ACKNOWLEDGEMENTS

First, I am very grateful to the commission board members of my thesis who accept and willing to read my thesis. I would like to thank two of my main supervisors, Prof. Catherine Corbier from France, and Associate Prof. Dr Razip Samian from Malaysia, who spent a lot of time reading my thesis and gave me a lot of useful suggestions and criticisms, and also for being patient with me through all the times. I really appreciate the guidance and assistances that they have provided and very grateful them to accept me as their student under the UHP-USM cotutelle program. I also like to express my gratitude to Dr Claude Didierjean who shared with me a lot of his expertise and research insight. He quickly became for me the role model of a successful researcher in the field. I could not have imagined having a better mentor, and without his enthusiasm, his inspiration, his unselfish passion and his encouragement, I would never have finished.

The members of my supervisory committee provided with many useful suggestions and comments to me. This includes Prof. Omar Shawkataly, Associate Prof. Dr. Sheila Nathan and Dr. André Aubry. I want to thank Dr. Fréderique Favier and Dr. Guillermo Mulliert, who taught me the knowledge of protein X-ray crystallography and protein modeling, which they have provided me a much broader understanding of these subjects. Many thanks go to the director of LCM^{3}B Prof. Claude Lecomte who welcomed me in the laboratory and to the present and past LCM^{3}B members particularly Dr. Moutsé Ranaivoson, Sébastien Moniot, Dr. Hélène Dubourg, Dr. Fabien Pascale, Dr. Sébastien Lebègue, Sandrine Mathiot, Alexandre Bouché, Dorothee Liebschner, Mikael Elias etc. who offered countless aides, discussions and provided a friendly working environment in the laboratory.

Special thanks to the team of Prof. Jean-Pierre Jacquot, particularly Prof. Jacquot himself, Dr. Nicolas Rouhier, Dr. Eric Gelhaye, Dr. Nicolas Navrot and Dr. Alison Winger, who provided proteins supply generously and had given me some useful suggestions about this project from biochemical aspects.

The financial support of the Universiti Sains Malaysia (ASTS fellowship) and the French Government (financial aids from EGIDE, CUGN and CNRS) is gratefully acknowledged. I would like to acknowledge the FIP-team in ESRF, Grenoble, France and the DESY-team in EMBL-Hamburg Outstation, Germany, for providing access to synchrotorn beamlines.

I would also like to give special thanks to Nicolas who had given me uncountable useful advices and have comforted me when I faced some setbacks in my research work. Thank you for sharing your opinions with me and I am glad to have a best friend like you. I will remember you forever and will not forget the good time we have together. Special thanks also dedicated to Moutsé and Alison who encourage me to be strong and optimistic in life. I also want to thank you my friends in Nancy who have given me tremendous spiritual support throughout my study in France; they are Thu, Cao Lei, Na Du, Andrew, Hana, Ezatul, Syukri and Nazeer not forgetting my friends in Malaysia; they are Bee Ben, Kay Keat, Kue Peng, Hooi Ling, Yu Ching and many more. Because the list is too long, I will simply say thank you very much to all of my friends which, from my childhood until graduate school, have joined me in the discovery of what is life about and how to make the best out of it.

Finally, and most importantly, I give my distinguished appreciation to my family. They have always supported and encouraged me throughout my postgraduate research life. To them I dedicate this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCD	Charge-Coupled Device
COX	Cytochrome c oxidase
DHAR	Dehydroascorbate reductase
DTT	Dithiothreitol
FT	Fourier Transform
FTR	Ferredoxin/Thioredoxin Reductase
Gpx	Glutathione peroxidase
GR	GSH-reductase
Grx	Glutaredoxin
GSH	Glutathione
GST	Glutathione-*S*-transferase
HED	β-hydroxyethyl disulfide
IR	Infrared
MAD	Multi-wavelength anomalous diffraction
MAPK	Map Kinase
MAPKKK	Map Kinase Kinase Kinase
MD	Molecular Dynamics
MIR	Multiple Isomorphous Replacement
MPD	2-Methyl-2,4-Pentanediol
NMR	Nuclear Magnetic Resonance
NTR	NADPH/ Thioredoxin reductase
OB fold	Oligonucleotide-binding fold

PDB	Protein Data Bank
PDI	Protein Disulfide Isomerase
PEG	Polyethylene glycol
PICOT	Protein Kinase C-interacting cousin of thioredoxin homology domain
Prx	Peroxiredoxin
ROS	Reactive Oxygen Species
SAD	Single-wavelength anomalous diffraction
SIRAS	Single isomorphous replacement with anomalous scattering
S*n*B	Shake-and-Bake
TLS	Translation, Libration, Screw
Trx	Thioredoxin
TrxR	Thioredoxin reductase
VMD	Visual Molecular Dynamics
WT	Wild-type

OKSIDOREDUKTASE DARIPADA KAYU KERAS JENIS POPULUS (POPLAR) YANG TERLIBAT DALAM RESPONS TERHADAP TEKANAN OKSIDATIF: KAJIAN KRISTALOGRAFIK KE ARAH PEMAHAMAN MEKANISME PEMANGKINAN ENZIM

ABSTRAK

Penghasilan species oksigen reaktif (ROS) merupakan akibat daripada persekitaran sentiasa dalam keaadaan aerobik yang tidak boleh dielakkan dan ia boleh disingkirkan oleh beberapa jenis sistem enzim sebaik sahaja dihasilkan. Tiga jenis oksidoreduktase (Glutathion peroksidase, Gpx; thioredoksin, Trx; and glutaredoksin, Grx) daripada *Populus trichocarpa* × *deltoides* (kayu keras jenis poplar) telah dicirikan melalui pendekatan biokimia dan kristalografi X-ray. Gpxs terdiri daripada sekumpulan enzim yang mengawal tahap ROS dalam sel dan melindunginya daripada kerosakan oksidatif. Dalam kajian ini, saya telah mencirikan struktur kristal bagi Gpx5 poplar (PtGpx5) dalam bentuk terturun dan teroksida. Perbandingan antara kedua-dua struktur redoks menunjukkan bahawa perubahan konformasi yang drastik adalah satu keperluan untuk menghampirkan dua residu sistin yang berjauhan supaya membentuk ikatan disulfida intra-molekul. Trxs ialah sekumpulan enzim yang mengawalatur pelbagai protein melalui proses penurunan thiol-disulfida. Objektif kajian ini adalah untuk menerangkan proses pemangkinan sejenis isoform Trx baru poplar, PtTrx*h*4, memandangkan enzim ini hanya boleh diturunkan oleh Grx dan bukannya oleh sistem NADPH:thioredoksin reduktase yang biasanya digunakan oleh Trx. PtTrx*h*4 mengandungi tiga sistin; satu di bahagian terminal N protein (Cys4) dan dua lagi (Cys58 and Cys61) di tapak aktif tipikal Trx (WC_1GPC_2). Analisis dua struktur kristal PtTrx*h*4 dalam kajian ini, jenis liar dan mutan C61S, membolehkan kita mencadangkan satu mekanisme pemangkinan empat langkah yang juga disokong oleh kajian enzimatik. Grxs pula adalah protein redoks yang menggunakan elektron daripada GSH untuk pemangkinan proses penukaran thiol-disulfida. Di sini, saya mencirikan struktur PtGrxS12 yang berkompleks dengan GSH, di mana ia adalah struktur pertama dari kumpulan Grx tumbuhan jenis sub-kelas 1 yang memiliki tapak aktif atipikal $_{28}WCSYS_{32}$. Protein ini mempunyai sistin tambahan (Cys87) dan peranan residu tersebut masih tidak diketahui. Maklumat tentang tapak ikatan GSH juga dirumuskan dalam kajian ini. Semua struktur protein yang dicirikan di sini mendalami pengetahuan kita tentang pemangkinan redoks dalam sel umbuhan dan interaksi enzim-substrat.

POPLAR OXIDOREDUCTASES INVOLVED IN THE OXIDATIVE STRESS RESPONSE: A CRYSTALLOGRAPHIC SNAPSHOT TOWARDS THE UNDERSTANDING OF THE CATALYTIC MECHANISMS

ABSTRACT

The production of reactive oxygen species (ROS) is an unavoidable consequence of living in an aerobic environment and once produced, it can be removed by several different enzyme systems. Three oxidoreductases (glutathione peroxidase, Gpx; thioredoxin, Trx and glutaredoxin, Grx) from *Populus trichocarpa* × *deltoides* (poplar tree) were characterized using biochemistry and X-ray crystallography approaches. Gpxs are a group of enzymes that regulate the levels of ROS in cells, and protect them against oxidative damage. In this study, I have determined the crystal structures of the reduced and oxidized form of poplar Gpx5 (PtGpx5). Comparison of both redox structures indicates that a drastic conformational change is necessary to bring the two distant cysteine residues together to form an intramolecular disulfide bond. Trxs are a family of ubiquitous enzymes which regulate various protein partners through the thiol-disulfide(s) reduction. The aim of this study is thus to precisely describe the catalytic mechanism of a new isoform of Trx that has been characterized in poplar, PtTrx*h*4, since it has been demonstrated recently to be reduced by Grx but not by the typical NADPH:thioredoxin reductase reducing system. PtTrx*h*4 contains three cysteines; one localized in an N-terminal extension (Cys4) and two (Cys58 and Cys61) in the usual Trx active site (WC_1GPC_2). Analyses of two PtTrx*h*4 crystal structures solved in this study, wild-type and C61S mutant, allow us to propose a four-step disulfide cascade catalytic mechanism in accordance with enzymatic studies. Grxs are highly conserved redox-proteins that utilize electrons from GSH particularly to catalyze thiol-disulfide exchange reactions. Here, I present the structure of glutathionylated PtGrxS12, the first structure of plant Grx of subclass 1 with an atypical $_{28}WCSYS_{32}$ active site. This protein possesses an additional cysteine (Cys87) in which the role of this extra active-site cysteine remains obscure. Details of the GSH binding site are summarized in this study. Protein structures solved here shed lights to our understanding of the redox mechanism in plants and to the enzyme-substrate interactions.

ÉTUDE CRISTALLOGRAPHIQUE DES OXYDORÉDUCTASES IMPLIQUÉES DANS LA RÉPONSE AU STRESS OXYDATIF CHEZ LE PEUPLIER POUR UNE MEILLEURE COMPRÉHENSION DES MÉCANISMES CATALYTIQUES

RÉSUMÉ

La production des espèces réactives de l'oxygène (ROS) est une conséquence inévitable de la vie dans un milieu aérobie. Une fois produites, ces espèces peuvent être éliminées par différents systèmes enzymatiques. Les structure de trois oxydoréductases (la glutathion peroxydase (Gpx), la thiorédoxine (Trx) et la glutarédoxine (Grx)) de *Populus trichocarpa* × *deltoides* (le peuplier) ont été résolues par diffraction des rayons X. Les Gpxs forment un groupe d'enzymes qui régulent la concentration des ROS dans les cellules, et qui les protègent des effets d'un stress oxydant. Contrairement à leurs homologues d'origine animale, les Gpxs végétales ne dépendent pas du glutathion (GSH) mais des Trx pour leur fonctionnement. De plus, elles possèdent une cystéine dans leur site actif au lieu d'une sélénocystéine. Dans cette étude, j'ai résolu les structures cristallines des formes réduite et oxydée de la Gpx5 de peuplier (PtGpx5) et montré que des changements conformationnels drastiques sont nécessaires pour permettre la formation d'un pont disulfure intramoléculaire. Les Trxs constituent une famille de petites protéines ubiquitaires, régulant diverses protéines cibles par la réduction de leur pont disulfure. Par cette étude structurale, mon objectif était de comprendre précisément le mécanisme catalytique d'une nouvelle isoforme caractérisée chez le peuplier, la PtTrx*h*4, dont la capacité à accepter des électrons de la Grx a été récemment démontrée. A la différence des autres Trxs *h* caractérisées jusqu'ici, la PtTrx*h*4 ne peut pas être réduite par la voie typique qui implique la NADPH-Trx réductase. Cette PtTrx*h*4 contient trois cystéines, la première localisée dans une extension en position N-terminale (Cys4) et deux (Cys58 et Cys61) situées dans le site actif classique (WC_1GPC_2) de la Trx. Les résolutions des structures de l'enzyme sauvage et du mutant C4S sous forme oxydée m'ont permis de proposer un mécanisme catalytique en quatre étapes en accord avec les études enzymatiques. Les Grxs sont des protéines qui utilisent des électrons de GSH en particulier pour catalyser des réactions d'échange de thiol-disulfure. Ici, je présente la structure de la PtGrxS12 (en complexe avec le GSH), la première structure de la Grx végétale sous-classe 1 avec un motif du site actif atypique $_{28}WCSYS_{32}$. Cette protéine possède une autre cystéine (Cys87), dont le rôle reste obscur. Le site de fixation du GSH est détaillé ultérieurement dans ce manuscrit. Ces structures éclaircissent notre compréhension du mécanisme d'oxydoréduction chez les végétaux et nous permettent de mieux comprendre les interactions enzyme-substrat.

Introduction

CHAPTER 1

INTRODUCTION

1.1 An overview of X-ray crystallography and protein crystallography

X-ray crystallography is the science of determining the arrangement of atoms within a crystal from the manner in which a beam of X-rays is scattered from the electrons within the crystal. The method produces a three-dimensional picture of the electron density within the crystal, from which the mean atomic positions, their chemical bond, their disorders and sundry other information can be derived. By definition, a crystal is a solid in which a particular arrangement of atoms (its unit cell) is repeated indefinitely along three principal directions known as the basis (or lattice) vectors, which are not necessarily perpendicular. A wide variety of materials can form crystals; such as salts, metals, minerals, semiconductors, as well as various inorganic, organic and biological molecules, which has made X-ray crystallography fundamental to many scientific fields.

The oldest and most precise method of X-ray crystallography is single-crystal X-ray diffraction, in which a beam of X-rays is reflected from evenly spaced planes of a single crystal, producing a diffraction pattern of spots called reflections. Each reflection corresponds to one set of evenly spaced planes within the crystal. The density of electrons within the crystal is determined from the position and intensity of the various reflections observed as the crystal is gradually rotated in the X-ray beam; this density allows the atomic positions to be inferred. For single crystals of sufficient purity and regularity, X-ray diffraction data can determine the mean chemical bond lengths and angles to within a few thousandths of an Angstrom and to

within a few tenths of a degree, respectively. The data also allow the static and dynamic disorder in the atomic positions to be estimated, which is usually less than a few tenths of an Angstrom. At its most basic level, X-ray crystallography is useful in identifying known materials, characterizing new materials and in discerning materials that appear similar by other experiments. However, X-ray crystal structures have many other applications; for example, they can account for unusual electronic or elastic properties of a material, shed light on chemical interactions and processes, or serve as the basis for understanding enzymatic mechanisms and designing inhibitors of therapeutic roles with hope of preventing diseases. The term "X-ray crystallography" is also sometimes applied to methods that involve X-ray diffraction from polycrystalline materials, such as powders of small crystals studied by X-ray powder diffraction.

Early protein crystallographers, proceeding by analogy with studies of other crystalline substances, examined dried protein crystals and obtained no diffraction patterns. Thus X-ray diffraction did not appear to be a promising tool for analyzing proteins. In 1934, J. D. Bernal and Dorothy Crowfoot (later Hodgkin) measured diffraction from pepsin crystals still in the mother liquor. Bernal and Crowfoot recorded sharp diffraction patterns, with reflections out to distances in reciprocal space that inversely correspond in real space to the distances between atoms. The announcement of their success was the birth announcement of protein crystallography. Ever since, X-ray crystallography has been the most prolific technique for the structural analysis of proteins and protein complexes, and is still the "gold standard" in terms of accuracy and resolution (Russell et al. 2004). Each new structure provided a specific step forward, each a harder puzzle due to its size, the resolution needed or any number of other complicating factors. To list but a few,

GroEL-GroES (Xu et al. 1997), F_1-ATPase (Abrahams et al. 1994, Leslie & Walker 2000), the nucleosome (Davey et al. 2002, Schalch et al. 2005), a bacterial potassium ion channel (Doyle et al. 1998), photosystem I (Ben-Shem et al. 2003, Amunts et al. 2007), the ribosome (Ban et al. 2000), the plasma membrane proton pump (Pedersen et al. 2007) and a bacterial multidrug ABC transporter (Dawson & Locher 2006), certainly make the milestones seem closer as the protein structures proliferate. Nevertheless, the number of structures of macromolecular assemblies solved by X-ray crystallography is still quite small compared to that of the individual proteins and it will probably be many years before we have a complete repertoire of high-resolution structures for the hundreds of complexes in a typical cell. This discrepancy is due mainly to the difficult production of sufficient quantities of the sample and its crystallization (Russell et al. 2004).

1.1.1 Nobel prize winners associated with crystallography

The beauty and regularity of crystals impressed people such an extent that, in the past, crystals were regarded as products of nature with mysterious properties. Scientific investigation of crystals started in 1669, when Nicolaus Steno, a Dane working as a court physician in Tuscan, proposed that during crystal growth, the angles between the faces remained constant. For a given crystal form, individual crystals might differ in shape (for example, in the development of their faces), but they always have identical angles between the same faces.

Before the famous first X-ray crystallographic diffraction experiment by von Laue, Friedrich and Knipping in 1912, the internal regularity of a crystal was suggested but never proven. X-ray crystallography has dramatically changed this

situation. The idea that crystals could be used as a diffraction grating for X-rays arose in 1912 in a conversation between Paul Peter Ewald and Max von Laue in Munich. Ewald had proposed a resonator model of crystals for his thesis, but this model could not be validated using visible light, since the wavelength was much larger than the spacing between the resonators. Von Laue realized that electromagnetic radiation of a shorter wavelength was needed to observe such small spacings, and suggested that X-rays might have a wavelength comparable to the unit-cell spacing in crystals. Working with Friedrich and Knipping, they shined a beam of X-rays through a sphalerite crystal and recorded its diffraction (a large number of well-defined spots arranged in a pattern of intersecting circles around the spot produced by the central beam) on a photographic plate. Von Laue developed a law that connects the scattering angles and the size and orientation of the unit-cell spacings in the crystal, for which he was awarded the Nobel Prize in Physics in 1914. Table 1.1 summarized the Nobel laureates with their contributions in the domain of crystallography.

Crystal structures of proteins began to be solved in the late 1950's, beginning with the structure of sperm whale myoglobin (Kendrew et al. 1960) and hemoglobin (Perutz et al. 1960) by Max Perutz and Sir John Cowdery Kendrew, for which they were awarded the Nobel Prize in Chemistry in 1962. Since then, the progress of protein crystallography has followed an evolutionary rather than a revolutionary path. Many important advances have been achieved on the way, but most of the methods used by Perutz are still valid today and still useful, albeit often in modified versions (Dauter 2005). Since that success, over 50000 X-ray crystal structures of proteins, nucleic acids and other biological molecules have been determined (according to PDB statistics in June 2008).

Table 1.1 Nobel Prize laureates in scientific fields associated with crystallography.

Year	Domain	Nobel Laureate(s)	Findings / Scientific contributions
1901	Physics	W.C.Röntgen	Discovery of X-rays
1914	Physics	M Von Laue	Diffraction of X-rays by crystals
1915	Physics	W.H.Bragg & W.L.Bragg	Use of X-rays to determine crystal structure
1929	Physics	L-V de Broglie	The wave nature of the electron
1937	Physics	C.J.Davisson & G.Thompson	Diffraction of electrons by crystals
1946	Chemistry	J.B.Sumner	Discovery of enzymes that can be crystallized
1954	Chemistry	L.C.Pauling	Research of the nature of the chemical bond and its application to the elucidation of the structure of complex substances
1962	Physiology or Medicine	F.Crick, J.Watson & M.Wilkins	Discovery of the helical structure of DNA Books about the double helix
1962	Chemistry	J.C.Kendrew & M.Perutz	Studies of the structures of globular proteins
1964	Chemistry	D.Hodgkin	Determination of many biochemical substances structures including Vitamin B12
1976	Chemistry	W.N.Lipscomb	Elucidation of the structure of boranes
1982	Chemistry	A.Klug	Development of crystallographic electron microscopy and discovery of the structure of biologically important nucleic acid-protein complexes
1982	Physics	K.G.Wilson	Conceptualizing the theory of critical phenomena in connection with phase transitions
1985	Chemistry	H.Hauptmanm & J.Karle	Development of direct methods for the determination of crystal structures
1988	Chemistry	J.Deisenhofer, R.Huber & H.Michel	Determination of the three-dimensional structure of a photosynthetic reaction centre
1991	Physics	P-G de Gennes	Development of the methods of discovering order in simple systems that can be applied to polymers and liquid crystals
1992	Physics	G.Charpak	Discovery of the multi wire proportional chamber
1994	Physics	C.Shull & N.Brockhouse	Neutron diffraction
1996	Chemistry	R.Curl, H.Kroto & R.Smalley	Discovery of the fullerene form of carbon
1997	Chemistry	P.D.Boyer, J.E.Walker & J.C.Skou	Elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP) and discovery of an ion-transporting enzyme
2002	Chemistry	J.B.Fenn, K.Tanaka & K.Wüthrich	Development of methods for identification and structure analyses of biological macromolecules
2002	Physics	R.Davis Jr., M.Koshiba & R.Giacconi	Pioneering contributions to astrophysics (detection of cosmic neutrinos and the discovery of cosmic X-ray sources)
2003	Chemistry	P.Agre & R.Mackinnon	Discoveries concerning (water and ion) channels in cell membranes
2006	Chemistry	R.D.Kornberg	Studies of the molecular basis of eukaryotic transcription

1.1.2 Integration and applications of crystallography in structural biology today

The development and application of technologies in structural biology (availability of complete genome sequences, automation of cDNA cloning, automated protein expression screens, affinity tags and parallel purification strategies, high-throughput crystallization procedures, usage of third-generation synchrotron, automated sample changing and the use of high-energy sources and CCD detector technology) do accelerate the transition from gene to structure (Dry et al. 2000). X-ray crystallography has benefited from several technological advances in recent years that make the genome-wide protein structure determination a practical reality. Indeed, it is now used routinely by scientists to determine how a pharmaceutical interacts with its protein target and what changes might be advisable to improve it (Scapin 2006). The number of protein structures deposited in the Protein Data Bank now (June 2008) is almost 48 000, with the vast majority (85 %) determined using X-ray crystallographic methods. Among the available protein crystal structures, only 34 % (~14000 structures) of them actually represent unique proteins (structures with similar sequences at 90 % identity were excluded). Thousands of studies describing such structures have been published in the scientific literature, and 14 Nobel prizes in chemistry or medicine have been awarded to protein crystallographers (see Table 1.1).

There is a wide spectrum of experimental and computational methods for the identification and structural characterization of macromolecular complexes. These methods need to be combined into hybrid approaches to achieve greater accuracy, coverage, resolution and efficiency than any of the individual methods. New methods must be capable of generating possible alternative models consistent with

information such as stoichiometry, interaction data, homology to known structures, docking results and low-resolution images. There is a need to describe the structures and dynamics of stable and transient complexes (Russell et al. 2004). As motions of macromolecules throughout their conformational landscapes generate biological activity, these motions can be investigated by NMR, neutron scattering, molecular dynamics, UV to IR spectroscopy, cryo-electron microscopy, and visible-ray or X-ray techniques (Bourgeois & Royant 2005). X-ray crystallography provides insight into protein dynamics *via* the analysis of mean-square atomic displacements and TLS (translation, libration, screw) parameters (Chaudhry et al. 2004), or when a series of static structures assigned to various states along a reaction pathway is morphed into movies (Echols et al. 2003).

Structure-based drug design has contributed to the discovery of a number of drugs and late-stage clinical candidates. It is now common for a series of ligand-protein structures to be available in discovery projects. The availability of X-ray derived structural information on protein-ligand complexes is increasing, and this is a useful tool in lead optimization (Davis et al. 2003). Indeed, knowledge of 3-dimensional structure based function properties of a drug target is very essential for a successful *in silico* designing of drugs (Kishan 2007). However, some difficulties during structure determination processes and lack of knowledge of conformational freedom associated with available protein structures often hurdle the structure based drug designs.

Structural genomics may contribute to a comprehension and efficient structural description of complexes in an additional way. Although structural genomics currently focus on single proteins or their domains, it could be expanded to the sampling of domain-domain interactions (Aloy et al. 2003, Apic et al. 2001, Sali

2003). Indeed, recent developments in the experimental and computational techniques have allowed structural biology to shift its focus from structures of individual proteins to the structures of large assemblies (Sali et al. 2003, Baumeister 2002, Sali & Kuriyan 1999). In contrast to structure determination of individual proteins, structural characterization of macromolecular assemblies usually poses a more difficult challenge. Nevertheless, a comprehensive structural description of large complexes generally requires the use of several experimental methods in order to maximize efficiency, completeness, accuracy and resolution (Sali et al. 2003). Coupled with X-ray crystallography, hybrid approaches have been successfully elucidating models of the actin-myosin complex (Pellecchia et al. 1999), the yeast ribosome (Spahn et al. 2001, Beckmann et al. 2001) and 20S proteasome (Groll & Huber 2005, Borissenko & Groll 2007), pre-mRNA splicing complex SF3b (Golas et al. 2003) and complex virus structures (Zhou et al. 2001, Baker et al. 2003).

Structural biology is a great unifying discipline of biology. Thus, structural characterization of many protein complexes may be the way to bridge the gaps between genome sequencing, functional genomics, proteomics and system biology. The goal seems daunting, but the prize will be commensurate with the effort invested, given the importance of molecular machines and functional networks in biology and medicine (Russell et al. 2004). It seems that in the near future, macromolecular crystallography will move along two parallel paths, system-oriented and discovery-oriented (Stevens 2004). The system-oriented approach corresponds to the traditional, slower, but more focused way of solving crystal structures of macromolecules and the discovery-oriented approach is, in other words, the structural genomics, speedier and wider track.

1.1.3 Understanding the basis of crystallography

When we see an object, light rays bounce off (are diffused by) the object and enter the eye through the lens, which reconstructs an image of the object and focuses it on the retina. In a simple microscope, an illuminated object is placed just beyond one focal point of a lens, which is called the objective lens. The lens collects lights diffused from the object and reconstructs an image beyond the focal point on the opposite side of the lens, as shown in Figure 1.1.

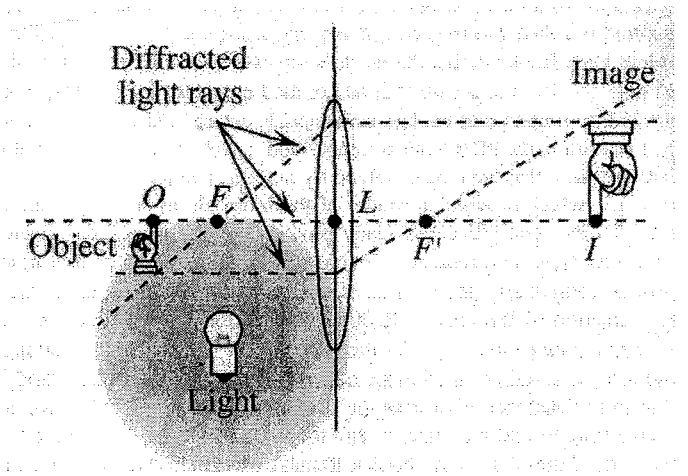

Figure 1.1 Action of a simple lens. Rays parallel to the lens axis strike the lens and are refracted into paths passing through a focus (*F* or *F'*). Rays passing through a focus strike the lens and are refracted into paths parallel to the lens axis. As a result, the lens produces an image at I of an object at *O* such that (*OF*) (*IF'*) = (*FL*) (*F'L*) (Rhodes 2006).

For a simple lens, the relationship of object position to image position in Figure 1.1 is (*OF*) (*IF'*) = (*FL*) (*F'L*). Because the distances, *FL* and *F'L* are constant (but not necessarily equal) for a fixed lens, the distance *OF* is inversely proportional to the distance *IF'*. Placing the object just beyond the focal point *F* results in a magnified image produced at a considerable distance from *F'* on the other side of the

lens, which is convenient for viewing. In a compound microscope, the most common type, an additional lens, the eyepiece, is added to magnify the image produced by the objective lens.

In order for the object to diffract light and thus be visible under magnification, the wavelength (λ) of the light must be, roughly speaking, no larger than the object. Visible light, which is electromagnetic radiation with wavelengths of 400 - 700 nm (nm = 10^{-9} m), cannot produce an image of individual atoms in protein molecules, in which bonded atoms are only about 0.15 nm or 1.5 angstroms (Å = 10^{-10} m) apart. Electromagnetic radiation of this wavelength falls into the X-ray range, so X-rays are diffused by even the smallest molecules.

Even though individual atoms diffract X-rays, it is still not possible to produce a focused image of a single molecule, for two reasons. First, X-rays cannot be focused by lenses. Crystallographers sidestep this problem by measuring the directions and strengths (intensities) of the diffracted X-rays and then using a computer to simulate an image-reconstructing lens. In short, the computer acts as the lens, computing the image of the object and then displaying it on a screen. Second, a single molecule is a very weak scatterer of X-rays. Most of the X-rays will pass through a single molecule without being diffracted, so the diffracted beams are too weak to be detected. Analyzing diffraction from crystals, rather than individual molecules, solves this problem. A crystal of a protein contains many ordered molecules in identical orientations, so each molecule diffracts identically, and the diffracted beams for all molecules augment each other to produce strong, detectable X-ray beams.

In brief, determining the structure of a protein by X-ray crystallography entails growing high-quality and well-ordered crystals (that will diffract X-rays

strongly) of the purified protein, measuring the directions and intensities of X-ray beams diffracted from the crystals (regular and repeating array of many identical molecules), and using a computer to simulate the effects of an objective lens and thus produce an image of the crystal's contents. Finally, the crystallographer must interpret that image, which entails displaying it by computer graphics and building a molecular model that is consistent with the image.

Electron densities can be probed by the scattering of X-rays. This is usually done by creating crystals, containing many trillions of molecules arranged in a regular three-dimensional pattern. When this pattern is illuminated with X-rays, diffraction of the X-rays occurs. The diffraction pattern encodes the information about the electron density distribution in the crystal. The diffracted rays are themselves, like the incident X-rays, electromagnetic waves with amplitude and phase and form a three-dimensional pattern, known as the reciprocal lattice. They are an indirect image of the electron density and are related to it by a Fourier transform. To recover the electron density from this diffraction pattern by inverse Fourier transform, both amplitudes and phases are needed. Experimentally usually only the amplitudes can be measured. The phase information remains hidden within these amplitude data and has to be recovered by computational techniques, often requiring additional measurements. This situation is known as the "Phase Problem in Crystallography". It is indeed a problem, because the phases influence the electron density to a much larger extent than the amplitudes do.

1.1.4 Protein crystals

1.1.4.1 Preparation and handling of biological macromolecules for crystallization

In the crystallization of biological macromolecules the quality (purity and homogeneity) and quantity of the required material is important. Difficulties in crystal growth sometimes may be linked to the nature or source of the biological material. Better crystallization conditions or diffracting crystal habits could be found by switching from one organism to another. This is because variability in sequences between heterologous species may lead to different conformations, and consequently to different crystallization behaviors. In practice, proteins isolated from eukaryotes are frequently more difficult to crystallize than their prokaryotic counterparts. Often their degree of structural complexity is higher. They can possess additional domains that may contribute to less compact and/or more flexible structures. Post-translational modifications (addition of other biochemical functional groups, changing the chemical nature of an amino acid or making structural changes like the formation of disulfide bridges) are often responsible for structural or conformational microheterogeneity. Purification, stabilization, storage and handling of macromolecules are therefore essential steps prior to crystallization. The "freshness" of the starting material and physiological state of cells is very important. Some proteins from unicellular organisms have been isolated in their native state only when cells are in exponential or pre-stationary growth phase (for an example, see Lorber & DeLucas 1990).

Special cares are needed in handling pure macromolecules in order to ensure that they are not damaged or lost before or during the crystallization trials and for

their reproducibility. To achieve this, several guidelines should be taken into account (see Ducruix & Giegé 1999 for details).

1. Diluted protein solutions are concentrated by ultrafiltration in devices using pressure (optimizing stir rate to prevent adsorption on to membrane surfaces and damage by shearing) or centrifugal force (remove aggregates formed as a result of a decrease in pH, of oxidation, or of an increase in salt or protein concentration), by dialysis against hygroscopic compounds (e.g. PEG or gel-filtration matrices), or by precipitation (addition of ammonium sulphate).
2. High concentrations of denaturing agents should be avoided because they inactivate or unfold proteins. In contrast, mild non-ionic detergents (e.g. octyl glucoside, heptyl thioglucoside and alkyl thiomaltosides are very useful in membrane protein biochemistry) help to solubilize proteins.
3. For better reproducibility, the pH of buffers after mixing of all ingredients should be adjusted since it may change after dilution or in the presence of other compounds.
4. Freeze-thawing of macromolecules should never be repeated to avoid denaturation. Therefore, experimentation on aliquots to limit repeated handling of stock solutions is highly recommended.
5. Any undesired molecules that might hinder crystallization should be removed by dialysis (e.g. glycerol, excess ligands), ultrafiltration, or size-exclusion chromatography.
6. It is also worthwhile to prepare macromolecules with or without their ligands (e.g. coenzyme, metal ions) or try additives (e.g. ions, reducing agents, chelators) because one or the other form may be more able to crystallize.

1.1.4.2 Properties of protein crystals

Whereas inorganic crystals can often be grown to dimensions of several centimeters or larger, it is frequently impossible to grow protein crystals as large as 1 mm in their shortest dimension. In addition, larger crystals are often twinned (two or more crystals grown into each other at different orientations) or otherwise imperfect and not usable. Roughly speaking, protein crystallography requires a crystal of at least 0.2 mm in its shortest dimension, although modern methods of data collection can sometimes succeed with smaller crystals, and modern software can sometimes decipher data from twinned crystals.

Inorganic crystals derive their structural integrity from the electrostatic attraction of fully charged ions. On the other hand, protein crystals are held together by weaker forces, primarily hydrogen bonds between hydrated protein surfaces. In other words, proteins in the crystal stick to each other primarily by hydrogen bonds through intervening water molecules. Protein crystals are thus much more fragile than inorganic crystals; gentle pressure with a needle is enough to crush the hardiest protein crystal. Growing, handling, and mounting crystals for analysis thus require very gentle techniques. If possible, protein crystals are often harvested, examined, and mounted for crystallography within their mother liquor, the solution in which they formed. This is because protein crystals are not only susceptible to chemical but also physical degradation, therefore they need to be maintained and stabilized in an equilibrated liquid environment as how they were formed and yet crystals will not be resolved (some sort of encapsulating effect).

Real macroscopic crystals are actually mosaics of many submicroscopic arrays in rough alignment with each other. The result of mosaicity is that an X-ray

reflection actually emerges from the crystal as a narrow cone rather than a perfectly linear beam. Thus the reflection must be measured over a very small range of angles, rather than at a single, well-defined angle. In protein crystals, composed as they are of relatively flexible molecules held together by weak forces, this mosaicity is more pronounced than in crystals of rigid organic or inorganic molecules, and the reflections from protein crystals therefore suffer greater mosaic spread than do those from more ordered crystals. The crystal mosaicity is a direct indicator of the physical perfection of the macromolecular crystal and it provides a simple measurement of crystal quality independent of many experimental parameters. It has been used to characterize successfully the improvement seen in some microgravity samples, for example, a reduction in the reflection mosaic spread providing a corresponding increase in the signal-to-noise ratio of the reflection (Snell et al. 1995, Ng et al. 1997).

Careful analysis of electron-density maps usually reveals many ordered water molecules on the surface of the protein in the structure. Additional disordered water is presumed to occupy regions of low density between the ordered particles. Ordered water molecules refer to water molecules that occupy the same site on every protein molecule in every unit cell (or a high percentage of them) and thus show up clearly in electron-density maps. Disordered water molecules refer to bulk water molecules that occupy the spaces between protein molecules which are in different arrangements in each unit cell, and thus show up only as uniform regions of low electron density. The quantity of water varies among proteins and even among different crystal forms of the same protein. The number of detectable ordered water molecules averages about one per amino-acid residue in the protein. Both the ordered and disordered water are essential to crystal integrity and maintaining the activity of

protein molecules in crystalline form (Frey 1994, Timasheff 1995), so drying destroys the crystal structure. For this reason, protein crystals are subjected to X-ray analysis in a very humid atmosphere or in a solution that will not dissolve them, such as the original mother liquor or a protective harvest buffer (with a higher concentration of precipitant or supplemented with cryoprotective agents such as PEG 400, PEG 600, glycerol or MPD) (Heras & Martin 2005).

NMR analysis of protein structure suggests that the ordered water molecules seen by X-ray diffraction on protein surfaces have very short residence times in solution. Thus most of these molecules may be of little importance to an understanding of protein function. However, ordered water molecules are of great importance to the crystallographer. As the structure determination progresses, ordered water molecules become visible in the electron-density map. Assignment of water molecules to isolated areas of electron density (small regions of disconnected density) improves the overall accuracy of the model, and improvements in accuracy in one area of the model give accompanying improvements in all other regions.

1.1.4.3 Evidence that solution and crystal structures are similar

Well-ordered crystals are difficult to grow because globular protein molecules are large, spherical, or ellipsoidal objects with irregular surfaces, and it is impossible to pack them into a crystal without forming large holes or channels between the individual molecules. These channels, which usually occupy more than half the volume of the crystal, are filled with disordered solvent molecules. The protein molecules are in contact with each other at only a few small regions, and even in these regions many interactions are indirect, through one or several layers of solvent

molecules. This is one reason why structures of proteins determined by X-ray crystallography are the same as those for the proteins in solution.

In a few cases, the structure of a protein has been obtained from more than one type of crystals (in different space groups). The resulting subunit models were identical, suggesting that the molecular structure was not altered by crystallization. One of the recent examples is the crystal structure of the Type IIP restriction endonuclease MspI bound to DNA (Xu et al. 2005) containing its cognate recognition sequence that has been determined in both monoclinic and orthorhombic space groups. These two independent crystal forms present an identical structure of a novel monomer-DNA complex, suggesting that this is not merely a crystallographic artifact.

1.1.4.4 Growing protein crystals

1.1.4.4.1 The crystallization phase diagram

The crystallization process can be illustrated by a phase diagram, which indicates which state (liquid, crystalline or amorphous solid [precipitate]) is stable under a variety of crystallization parameters. It provides a mean of quantifying the influence of certain parameters, such as the concentration of protein, precipitant(s), additive(s) and so on, on the production of crystals. Thus, phase diagrams form the basis of the design of crystal optimization experiments (Ataka 1993, McPherson 1999, Ducruix & Giegé 1999).

Figure 1.2 illustrates a typical crystallization phase diagram, consisting of four zones representing different degrees of supersaturation: a zone of high

supersaturation, where the protein will precipitate; a zone of moderate supersaturation, where spontaneous nucleation will take place; the metastable zone (just below the nucleation zone) of lower supersaturation, where crystals are stable and may grow but no further nucleation will take place (the conditions in this region are the best for the growth of well-ordered crystals); and a zone of undersaturation, where the protein is fully dissolved and will never crystallize (Chayen 2004, Ducruix & Giege 1999).

To obtain good structural data, crystals need to be single and have dimensions of at least 100 μm^3, preferably much larger. In an ideal experiment, once nuclei have formed, the concentration of protein in the solute will drop, thereby leading the system into the metastable zone, where few single crystals will grow (Figure 1.2). More often than not, either no crystals form at all or excess nucleation occurs, whereby numerous clusters of tiny crystals are formed instead of a few sizeable ones. An additional frustrating problem is the formation of large single crystals that do not diffract (Chayen 2004).

By examining the crystallization phase diagram and solubility properties of the sample, an understanding may be gained to optimize the process of crystal growth. The aim is to devise methods that will enable the experimenter to manipulate the phase diagram and actively control the crystallization environment in order to lead to crystal growth in the direction that will produce the desired results (Chayen 2005). This can be achieved in various ways (see section below), for example, control of the nucleation stage; dynamic separation of the nucleation and growth phases; and influencing the kinetics of the crystallization process.

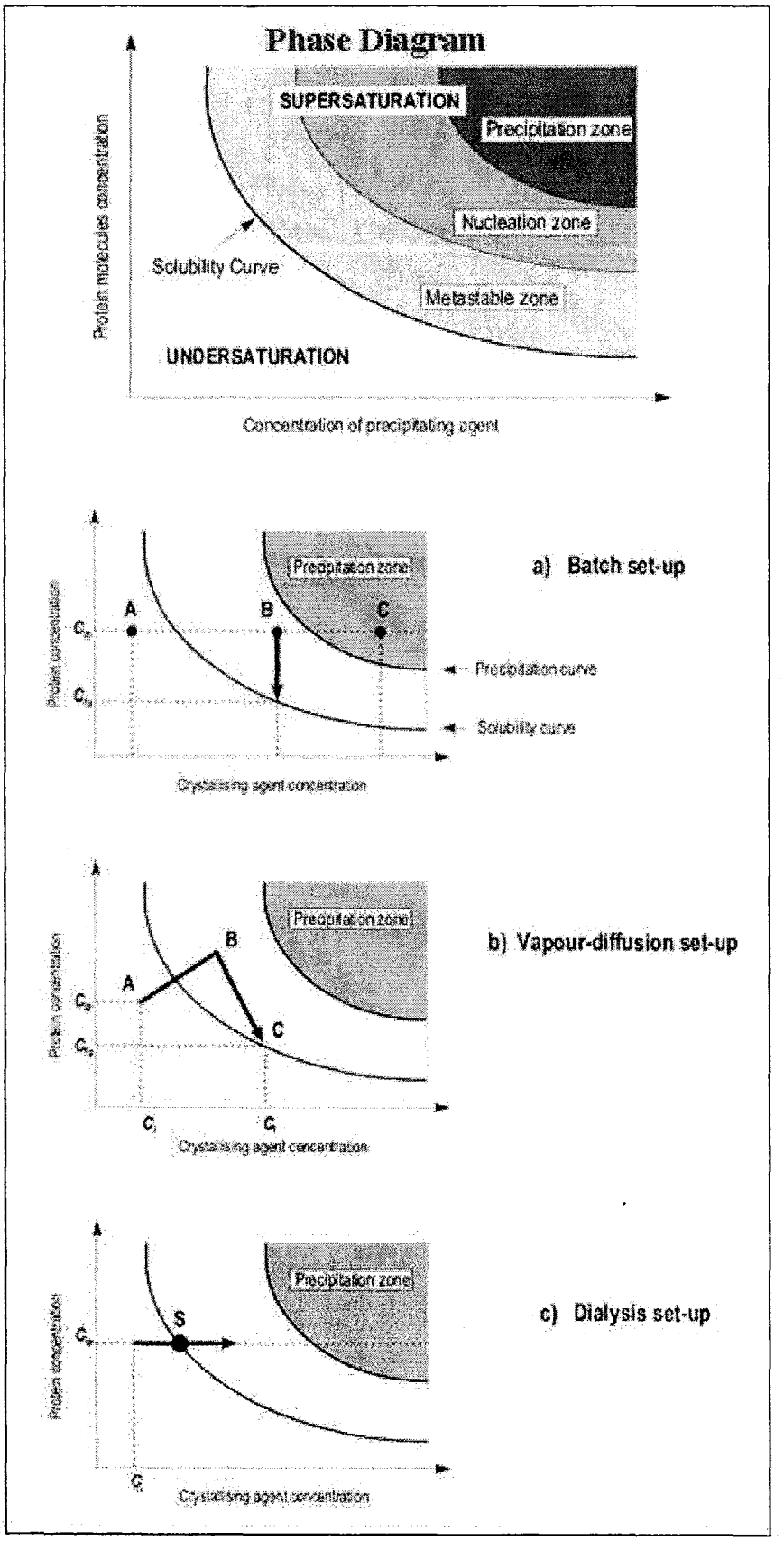
Phase Diagram
SUPERSATURATION
Precipitation zone
Nucleation zone
Solubility Curve
Metastable zone
UNDERSATURATION
Protein molecules concentration
Concentration of precipitating agent
Precipitation zone
A
B
C
Precipitation curve
Solubility curve
Protein concentration
Crystallising agent concentration
a) Batch set-up
Precipitation zone
A
B
C
Protein concentration
Crystallising agent concentration
b) Vapour-diffusion set-up
Precipitation zone
S
Protein concentration
Crystallising agent concentration
c) Dialysis set-up

Figure 1.2 Schematic illustration of a protein crystallization phase diagram (top panel) and various crystallization set-ups explained in terms of phase diagrams (bottom panel). The adjustable parameter can be precipitant or additive concentration, pH, temperature and so on. The three major crystallization methods are represented, showing their different routes to the nucleation and metastable zones, assuming the adjustable parameter is crystallizing agent concentration. The black circles represent the starting conditions. The solubility is defined as the concentration of the protein in the solute that is in equilibrium with crystals. The supersolubility curve is defined as the line separating conditions where spontaneous nucleation (or phase separation, precipitation) occurs from conditions where the crystallization solution remains clear if left undisturbed. In the bottom panel is the schematic representation of solubility phase diagram and correlation between protein and crystallizing agent concentrations in (a) batch, (b) vapour-diffusion and (c) dialysis crystallization experiments. C_{ip} and C_i are the initial concentrations of protein and crystallizing agent respectively, C_{fp} and C_f are their final concentrations (Figure adapted from Ducruix & Giegé 1999).

1.1.4.4.2 Factors influencing protein crystal growth

The two most important keys to success of a crystallographic project are purity and quantity of the macromolecule under study. Impure samples usually will not make suitable crystals, and even for proteins of the highest purity, repeated trials will be necessary before good crystals result.

Many variables influence the formation of protein crystals and each protein is unique. It is not possible to foresee the conditions that can cause or promote crystallization of a protein. The various parameters that affect crystallization are not independent of each other and their interrelation may be complicated and difficult to distinguish (McPherson 1999). There is no universal guideline to crystallize macromolecules successfully. The only way to do this is to identify the important components and refine each of them distinctively, based on knowledge on protein properties and its behavior observed from purification experiences. Again, these

components vary with proteins. In general, protein purity and concentration, precipitant type and concentration, buffer type and pH, and temperature are the most important factors for protein crystallization. They are normally being considered first when determine its effect on crystallization. More subtle ones like cleanliness, vibration and ultrasound (Edwards & Palmer 1990), convection, source and age of the protein and the presence of ligands are also among parameters that should also be taken into considerations. Clearly, the problem of developing a reliable source of crystals entails controlling and testing a large number of parameters. Indirectly, the difficulty and importance of obtaining good crystals has prompted the invention of crystallization robots that can be programmed to set up many trials under systematically varied conditions. However, when varying the more conventional parameters fails to produce good crystals, the crystallographer may take more drastic measures, e.g. mutagenesis. Sometimes limited digestion of the protein by a proteolytic enzyme removes a disordered surface loop, resulting in a more rigid, hydrophilic, or compact molecule that forms better crystals. A related measure is adding a ligand, such as a cofactor, that is known to bind tightly to the protein. The protein-ligand complex may be more likely to crystallize than the free protein, either because the complex is more rigid than the free protein or because the cofactor induces a conformational change that makes the protein more amenable to crystallizing. In a nutshell, Table 1.2 summarizes the factors (physical, chemical and biochemical factors) which affect the crystallization of macromolecules.

Table 1.2 Important factors affecting macromolecular crystallization (McPherson 1990; McPherson 1999).

Physical	Chemical	Biochemical
Temperature / temperature fluctuation	Buffer pH	Purity of macromolecule
Vibration / sound / mechanical perturbation	Precipitant type	Substrate / coenzyme / ligand / inhibitor / effectors
Time / rate of growth	Precipitant concentration	Inherent symmetric of the macromolecule
Equilibrium rate	Macromolecule concentration	Biochemical modification
Dielectric constant of medium	Ionic strength	Genetic / post-transitional modification
Viscosity of medium	Additive / specific ions	Isoelectric point
Pressure	Metal ions	Macromolecule stability
Gravity	Detergent / surfactant	Aggregation state of macromolecule
Homogeneity of macromolecule	Degree of supersaturation	Storage time of macromolecule
Electric / magnetic fields	Reducing / oxidizing environment	Source of macromolecule / history of sample
Volume of crystallization sample drop	Present of amorphous substances / impurities	Proteolysis / hydrolysis
Methodology / approach of crystallization	Cross-linker	Microorganism contamination

1.1.4.4.3 Crystallization techniques

1.1.4.4.3.1 Batch crystallization

The batch technique is attractive because of its inherent simplicity and reproducibility. It requires nothing more than the combination of two or more solutions (precipitating reagents and protein solution) into one, and a period of time until spontaneous nucleation commences, suddenly bringing the solution to a state of higher supersaturation. With luck, crystals grow gradually from the supersaturated solution without further processing. An automated system for microbatch

crystallization has been designed by Chayen et al. (1990, 1992). In the micro-batch technique (Figure 1.3a), smaller volumes, as little as 0.5 μl, can be used. But usually people tend to grow protein crystals in 2 to 3 μl drop, containing the protein and the precipitant, which is dispensed into the well of a "Terazaki-type" microtiter plate, covered with paraffin oil. The oil acts as a sealant to prevent evaporation. It does not interfere with the common precipitants, but it does interfere with organic compounds that dissolve in the oil (Chayen 1997, Chayen 1998). During the incubation period, the concentration of a precipitant agent remains constant since evaporation is limited and, therefore, the volume of the drop remains the same during the experiment. On the other hand, the concentration of the protein changes on formation of either crystals or amorphous precipitant. If the concentration of precipitant agent is chosen in such a way that the solution is in an undersaturated state, crystallization will never occur.

The main disadvantage of this method could be that the equilibration occurs very rapidly, thus affecting the rate of crystal growth (increases nucleation rate) and consequently decreases the size and the quality of crystals obtained. The manipulation of the crystals from the drop covered by oil could be difficult too. However, since the use of very small volumes (up to 50 nanoliters by using robotic system) of protein solution can be made, the micro-batch technique is quite useful as an initial screening method. Although the evaporation of water from the drop covered by oil is negligible, it does occur, and therefore the 'life-time' of micro-batch trials is usually about 2 to 3 weeks (see http://www-cryst.bioc.cam.ac.uk/ for details).