

Biomolecular Crystallography:

Principles, Practice and Application to Structural Biology

Bernhard Rupp

Part I: From sequence to crystals

Chapter 1 Introduction: Planning Your Study

- A brief overview of protein crystallography and its method and challenges preparing the ground for the subsequent in-depth chapters.

1. Introduction: Preparing for your study	1
1.1. Molecular structure defines function	2
1.2. Principle and challenges of crystallography: A preview	7
1.3. Protein molecules and the crystalline state	10
1.4. Interlude: Molecular cryptanalysis	12
1.5. Planning and executing a protein structure determination	15
1.5.1. Project planning and alternatives	16
1.5.2. Keeping track of the goal	17
1.5.3. Automation and parallel approaches	18
1.5.4. Protein engineering	19
1.5.5. Crystal growth	19
1.5.6. Harvesting, cryo-protection and mounting of crystals	20
1.5.7. Diffraction data collection	21
1.5.8. Phase determination	23
1.5.9. Electron density interpretation and model building	24
1.5.10. Restrained reciprocal space refinement	25
1.5.11. Structure validation	26
1.5.12. Analysis and description of the structure	26
1.6. Crystallographic models and coordinate files	28
1.6.1. The Protein Data Bank	28
1.6.2. Format of protein structure coordinate files	29
1.6.3. Graphical display of structure models	29
1.7. Crystallographic computer programs	30
1.8. Key concepts	31
1.9. Thought questions	32
1.10. Additional reading	32
1.11. References	33

Chapter 2. Protein structure

- Review of basic protein properties and nucleic acids with emphasis on needs of structural biologist and protein crystallographer.

2. Protein Structure	1
2.1. Introduction	3

2.2. Basics of protein structure	4
2.2.1. Fundamentals of molecular geometry	8
2.2.1.1. Bond distances, bond angles, and torsion angles	8
2.2.1.2. Chiral centers	10
2.3. The geometry of the polypeptide chain	11
2.3.1. Peptide bond basics and the protein backbone	11
2.3.2. <i>Trans</i> - and <i>cis</i> -peptide bonds	13
2.3.3. Backbone torsion angles - the Ramachandran plot	14
2.4. Secondary structure elements	16
2.4.1. Right handed α -helices	18
2.4.2. The 3_{10} -helix, the π -helix and PP-II helix	20
2.4.3. β -strand based sheet structures	22
2.4.4. β -bulges and irregular sheets	24
2.4.5. Turns and loops	25
2.4.5.1. Turns	25
2.4.5.2. Loops	26
2.5. Amino acids	28
2.5.1. Classification of residues	28
2.5.2. Hydrophobic residues	31
2.5.2.1. Hydrophobic core propensity	31
2.5.2.2. Aromatic phenylalanine	32
2.5.2.3. Proline	32
2.5.2.4. Methionine	33
2.5.2.5. Glycine	33
2.5.3. Charged residues	34
2.5.3.1. Acidic residues	34
2.5.3.2. Basic residues	35
2.5.4. Polar residues	36
2.5.4.1. Hydroxyl residues	36
2.5.4.2. Amide residues	36
2.5.4.3. Cysteine and disulfide bridges	37
2.5.5. Ambiguities in side chain electron density - NQH flips	38
2.5.6. Side chain torsions and multiple side chain conformations	39
2.5.7. Modified residues	41
2.5.8. About pKa values of amino acids	42
2.6. The interactions of side chains	43
2.6.1. Covalent interactions	44
2.6.2. Ionic interactions	45
2.6.3. Hydrogen bonds	46
2.6.4. Polar interactions	47
2.6.5. Van der Waals interactions	47
2.6.6. Hydrophobic interactions	48
2.6.7. Solvent structure	49
2.6.8. Hydration shells and water dynamics	49

2.7. Protein tertiary structure	50
2.7.1. Structural domains	51
2.7.1.1. Multi-domain proteins	51
2.7.1.2. Domain expression and truncation	51
2.7.2. Structural motifs	52
2.7.2.1. The β - α - β motif	52
2.7.2.2. The Calcium-binding EF-hand	53
2.7.3. Domain structure databases	54
2.7.3.1. Structure is more conserved than sequence	54
2.7.3.2. Major domain structure databases	55
2.7.4. Examples of domain architecture	55
2.8. Membrane proteins	58
2.8.1. Classification of membrane proteins	58
2.8.2. Polytopic membrane proteins are difficult to crystallize	60
2.9. Quaternary structure generates functional variety	62
2.9.1. Multimeric assemblies	62
2.9.2. Biological unit and protein quaternary structure	63
2.10. Nucleic acids and their protein complexes	65
2.10.1. Nucleic acid structure	67
2.10.2. Protein-DNA interactions	70
2.10.3. DNA recognition motifs	71
2.10.4. Nucleotide analogs and drugs	74
2.10.5. Organization of DNA in chromatin	76
2.11. Review of key concepts	78
Protein structure	78
DNA structure	79
Practical crystallography	80
2.12. Exercises	80
2.13. General references	81
2.14. Specific references	81

Chapter 3. Protein crystallization

- Develops theoretical concepts and practical challenges of protein crystallization

3. Protein crystallization	1
3.1. Overview of protein crystallization experiments	3
3.2. Understanding protein crystals	7
3.2.1. Properties and assembly of protein crystals	9
3.2.2. Crystal stability and intermolecular contacts	11
3.2.3. Local nature of crystal contacts	12
3.2.4. Crystal packing effects, artifacts, and solvent	13
3.2.5. Crystal forms and morphology	13

3.3. Protein properties and crystallization	14
3.3.1. Protein constructs and affinity tags.....	15
3.3.2. Batch variation and contaminants.....	16
3.3.3. Protein concentration	16
3.3.4. Purity, freshness, and conformational state	16
3.3.5. Buffers, salts and additives in protein stock.....	17
3.4. Thermodynamics of protein solutions.....	18
3.4.1. Solubility diagrams and supersaturation	18
3.4.2. Effect of pH on protein solubility.....	20
3.4.3. Effect of temperature on protein solubility.....	21
3.4.4. Entropic contributions drive crystallization.....	21
3.5. Self-assembly and crystallization kinetics.....	24
3.5.1. Activation and phase separation	24
3.5.2. Supersaturated solutions and nucleation.....	24
3.5.3. Crystallization diagrams.....	26
3.5.4. The growth of real crystals	28
3.6. Crystallization cocktails.....	31
3.6.1. Salts	32
3.6.2. Organic precipitants.....	33
3.6.3. Additives, detergents and cofactors.....	36
3.6.4. Buffers and the effect of pH on protein crystallization	38
3.7. Crystallization techniques.....	41
3.7.1. Vapor diffusion techniques.....	42
3.7.1.1. Hanging drop vapor diffusion	42
3.7.1.2. Crystallization diagram of a vapor diffusion experiment	43
3.7.1.3. Sitting drop vapor diffusion	44
3.7.2. Batch crystallization.....	46
3.7.3. Dialysis	48
3.7.4. Free-interface diffusion	49
3.7.5. Overlap between crystallization techniques	54
3.7.6. Control of nucleation by seeding.....	56
3.7.7. Robotics in the crystallization laboratory	58
3.8. Protein crystallization strategies	60
3.8.1. Against all odds	60
3.8.2. Crystallization as a multivariate sampling problem.....	61
3.8.3. Crystallization screen designs	62
3.8.4. Random sampling, reagent propensities, and number of experiments.....	63
3.8.5. Two-tiered approaches.....	66
3.8.6. Analyzing the outcome of crystallization trials.....	66
3.8.7. Optimization	71
3.9. Soaking and co-crystallization.....	73
3.9.1. Protein-ligand complexes	73
3.9.2. Binding constant and site occupation	74
3.9.3. Heavy atom soaking.....	77

3.9.4. Harvesting and mounting of crystals	80
3.10. Advanced crystallization problems	81
3.10.1. Integral membrane proteins	81
3.10.1.1. Detergent exchange and crystallization	83
3.10.1.2. Lipid cubic phases	84
3.10.2. Antibody scaffolding and F _{ab} structure	85
3.10.3. Protein-DNA complex crystallization	88
3.11. Key concepts	90
Fundamentals of crystallization	90
Crystallization techniques	92
Crystallization strategies	93
Special challenges	94
3.12. Thought questions	95
3.13. Additional reading	95
3.14. References	95

Chapter 4. Protein Production and Purification

- After having established that practically all crystallization trouble is rooted in protein properties, we can understand the specific needs for protein engineering for crystallography. This is reversed from actual workflow on purpose.

4. Proteins for crystallography	1
4.1. Overview of protein production	3
4.2. Engineering proteins that crystallize	5
4.2.1. Protein engineering strategies and levels	6
4.3. Rational design strategies at the DNA level	7
4.3.1. Bioinformatics tools	8
4.3.1.1. Sequence retrieval and alignment	8
4.3.1.2. Secondary structure, transmembrane- and disorder prediction	9
4.3.1.3. Analysis of sequence alignments	10
4.3.1.4. Domain expression and domain truncation	11
4.3.1.5. Active site predictions	12
4.3.2. Estimation of crystallization success	13
4.3.2.1. Predictions based on protein properties	13
4.3.2.2. Surface entropy reduction	16
4.3.2.3. Additional targeted, structure-based design strategies	18
4.3.3. Primer-based manipulation of target DNA	19
4.4. Combinatorial designs at the DNA level	20
4.4.1. Random DNA truncation and fragmentation	21
4.4.2. Directed evolution	22
4.5. Challenges in protein expression and purification	24

4.5.1. Advanced cloning techniques	25
4.5.2. Choice of expression system.....	26
4.5.2.1. Natural sources.....	27
4.5.2.2. Bacterial expression systems.....	27
4.5.2.3. Expression in yeasts.....	29
4.5.2.4. Insect and mammalian expression systems	30
4.5.2.5. Cell-free expression systems and chaperones	31
4.5.3. Fusions and tags for expression.....	33
4.5.3.1. Small peptide tags	33
4.5.3.2. Larger fusion tags and reporter systems.....	35
4.5.3.3. Coexpression	36
4.5.3.4. General remarks regarding protein purification	36
4.5.4. Labeling of proteins, Se-Met.....	38
4.5.5. Membrane protein expression and purification	41
4.5.5.1. The expression challenge	41
4.5.5.2. Solubilization and purification.....	42
4.6. Engineering strategies at the protein level.....	44
4.6.1. Limited proteolysis.....	44
4.6.2. Lysine methylation.....	45
4.6.3. Glycoproteins	46
4.7. Analysis of protein stability and conformational state.....	48
4.7.1. Thermofluor stability assays	48
4.7.2. Native PAGE and size exclusion chromatography	49
4.7.3. Light scattering methods	50
4.7.3.1. Static light scattering and molecular interactions.....	51
4.7.3.2. Dynamic light scattering and size distribution	53
4.7.4. Small angle X-ray and neutron scattering.....	55
4.7.5. Circular Dichroism spectroscopy and secondary structure	57
4.7.6. Nuclear magnetic resonance spectroscopy	58
4.7.7. Mass spectroscopic methods.....	60
4.8. Next steps	62
4.9. Key concepts	62
General protein engineering and design strategies	62
Predicting protein crystallizability	62
Combinatorial libraries	63
Protein expression	63
Protein labeling and protein modifications	64
Stability and conformational analysis	65
4.10. Thought questions	65
4.11. Additional reading.....	66
4.12. References	66

Part II: Fundamentals of protein crystallography**Chapter 5. Crystal geometry**

- The basic concepts of crystal geometry are developed starting from 2-dimensional examples, rapidly demonstrating the first important concepts, which are revisited in necessary detail during the remainder of the chapter.

5. Crystal geometry	1
5.1. Basic concepts of crystal assembly	3
5.1.1. Periodic lattices	4
5.1.2. Periodic patterns, motifs and unit cells	6
5.1.3. Alternate unit cell origins	8
5.1.4. Symmetry within the unit cell	10
5.1.4.1. Twofold rotations	11
5.1.4.2. Multiple unit cell choices	12
5.1.4.3. Four-fold rotations	13
5.1.4.4. The asymmetric unit of the unit cell	15
5.1.4.5. Three-fold and six-fold rotations	16
5.1.5. Limitations due to asymmetry of the motif	19
5.1.6. Non-crystallographic symmetry	22
5.1.7. NCS generated superstructures	26
5.2. Symmetry in three dimensions	28
5.2.1. Nomenclature of 3-dimensional lattices and unit cells	29
5.2.1.1. Lattice points, lattice vectors and fractional coordinates	31
5.2.1.2. Lattice translations and equivalent points	32
5.2.2. 3-dimensional crystal lattices	33
5.2.2.1. Primitive lattices	33
5.2.2.2. Centered lattices	34
5.2.3. Symmetry operations in 3-dimensional space	37
5.2.3.1. Screw axes	38
5.3. Space groups	42
5.3.1. Symmetry operators	43
5.3.2. Combination of symmetry operators	45
5.3.2.1. Point groups	46
5.3.2.2. Space group generators	46
5.3.2.3. Space group symbols	48
5.3.2.4. Space group tables	49
5.3.2.5. Generation of unit cell contents	53
5.3.2.6. Cartesian world coordinates	55
5.3.3. Space group preferences	59
5.4. The reciprocal lattice	61
5.4.1. Lattice planes	62
5.4.1.1. Two-dimensional lattices	62

5.4.1.2. 3-dimensional lattices and Miller indices	64
5.4.2. Construction of the reciprocal lattice	65
5.4.3. Centrosymmetry of the reciprocal lattice	69
5.5. Next steps	72
5.6. Key concepts	72
Fundamentals of crystal geometry	72
Basic crystal symmetry.....	73
Reciprocal lattice geometry.....	74
5.7. Thought questions	74
5.8. Additional reading.....	75
5.9. References	75

Chapter 6. Diffraction basics

- The previous chapter about crystal geometry and reciprocal space has introduced the basic concepts of symmetry and reciprocal space which are applied to principles X-ray scattering and diffraction, including anomalous scattering.

6. Diffraction basics	1
6.1. Scattering of X-rays.....	3
6.1.1. The nature of X-rays and their interaction with matter.....	5
6.1.2. Superposition of electromagnetic waves	9
6.1.3. Scattering diagrams	15
6.1.4. Scattering of X-rays by free electrons	16
6.1.5. Scattering by electrons in atoms	18
6.1.6. Atomic scattering factors	22
6.1.7. Debye-Waller factor, atomic displacement, and <i>B</i> -factor	25
6.1.8. Scattering from adjacent atoms	27
6.1.9. The Bragg equation	29
6.1.10. Scattering from a molecule	31
6.1.11. Diffraction from a molecular crystal	32
6.1.12. The structure factor.....	34
6.2. Diffraction geometry	37
6.2.1. The Ewald construction	37
6.2.2. Ewald sphere and reciprocal lattice	39
6.2.3. Diffraction limits and excluded regions.....	41
6.3. The intensity of diffracted X-rays	45
6.3.1. Structure factor and structure amplitude	45
6.3.2. Occupation and B-factor revisited.....	47
6.3.3. The first structure factor calculation.....	49
6.3.4. Structure factor calculation for a protein	51
6.3.5. Structure amplitude and intensity of diffraction	55
6.4. X-ray absorption	58
6.4.1. Variation of absorption with photon energy	58

6.4.2. X-ray absorption edges.....	60
6.4.3. Anomalous scattering factors.....	63
6.4.4. X-ray absorption spectra	64
6.4.5. Selection of wavelengths.....	66
6.4.6. Friedel pairs	68
6.4.7. The breakdown of Friedel's law	69
6.4.8. Selection of anomalous scatterers	70
6.4.9. Estimating the anomalous signal.....	73
6.4.10. Anomalous structure factor calculation for a protein	74
6.5. Reciprocal space symmetry revisited	76
6.5.1. Laue symmetry	77
6.5.2. Symmetry relations and Friedel pairs of the asymmetric reciprocal unit.....	81
6.5.3. Bijvoet pairs and absolute configuration	83
6.5.4. Centric and phase restricted reflections	85
6.5.5. Systematic absences.....	90
6.5.6. General position multiplicity and ϵ -factors	91
6.6. Review of key concepts	93
Scattering of X-rays.....	93
Structure factors	93
Diffraction of X-rays.....	95
Absorption of X-rays	95
Reciprocal space symmetry.....	96
6.7. General references	97
6.8. Specific references.....	97

Chapter 7. Statistics and probabilities in crystallography

- Many applications in crystallography are based on basic and advanced statistical concepts which are usually not taught - at least not at the right time and in suitable context - in biology-heavy curricula. An introduction to statistical methods and a primer in maximum likelihood and Bayesian inference as used in macromolecular crystallography is provided.

7. Statistics and probability in crystallography.....	1
7.1. Crystallography: applied probability theory	3
7.2. Probability distributions.....	4
7.3. Basic probability algebra.....	6
7.4. Photon counting statistics and Poisson distribution	10
7.4.1. Absolute and relative error.....	12
7.5. The normal distribution	13
7.5.1. Cumulative probability distributions	14
7.5.2. Accuracy vs. precision	16
7.5.3. Sampling and the central limit theorem.....	17

7.6. Descriptive statistics used in crystallography	18
7.7. Integration of probability distributions	22
7.8. Covariance and basic error propagation	23
7.8.1. Converting intensities to structure amplitudes	24
7.8.2. <i>R</i> -values and correlation coefficients	26
7.8.3. Weighted averages and merging of observations	27
7.9. Likelihood and Bayesian inference	29
7.9.1. Deductive vs. inductive reasoning	29
7.9.2. Bayes' theorem	31
7.9.3. Maximum Likelihood	34
7.9.3.1. Maximum Posterior versus Maximum Likelihood	34
7.9.3.2. Crystallographic models and conditioning information	35
7.9.3.3. Log-likelihood	36
7.9.3.4. Approximation of log-likelihood functions	38
7.9.3.5. Maximum likelihood principle, least squares and CLT	41
7.9.3.6. Multi-parametric likelihood functions	42
7.9.3.7. Optimization of crystallographic log-likelihood	43
7.9.4. Bayesian treatment of negative intensities	44
7.9.5. Are you crystallography-positive?	46
7.10. Unconditional structure factor probability distributions	48
7.10.1. The Wilson distribution for a random atom model	49
7.10.2. Normalized structure factors	55
7.10.3. Cumulative normalized intensity distributions	58
7.10.4. Mean $ E^2-1 $ value and expected random <i>R</i> -value	59
7.11. Conditional structure factor probabilities	60
7.11.1. Accounting for incomplete models - Sim weights	61
7.11.2. Accounting for errors in the model – anticipating σ_A	63
7.11.3. Improving the structure factor error model	64
7.11.4. Marginalizing the phase and the Rice distribution	67
7.11.5. Normalizing and analyzing the Rice and Woolfson distribution	69
7.12. Wilson plots and initial scaling	72
7.13. Review of key concepts	75
Probabilities and probability distributions	75
Common probability distributions and their properties	76
Error propagation	76
Intensity statistics and Wilson scaling	77
Bayesian inference and Maximum Likelihood	78
7.14. General references	79
7.15. Specific references	79

Part III: From crystal to data

Chapter 8. Instrumentation and data collection.

8. Instrumentation and data collection	1
8.1. X-ray instrumentation	4
8.1.1. X-ray sources	5
8.1.1.1. X-ray emission	6
8.1.1.2. Laboratory X-ray generators	8
8.1.1.3. X-ray optics	9
8.1.1.4. Synchrotrons	11
8.1.2. X-ray Area Detectors	16
8.1.2.1. Film and imaging plates	16
8.1.2.2. Multiwire array detectors	17
8.1.2.3. CCD detectors	18
8.1.3. Goniostats	21
8.1.4. Cryo-cooling equipment	23
8.1.5. Mounting robotics	24
8.2. Crystal harvesting and mounting	26
8.2.1. Harvesting and mounting of crystals	26
8.2.1.1. Capillary mounting	26
8.2.1.2. Cryomounting and cryoprotection	26
8.2.1.3. Advantages of cryocrystallography	29
8.2.1.4. Annealing and hyperquenching	31
8.2.2. Real crystals: mosaicity and twinning	33
8.2.2.1. Macroscopic and non-merohedral twinning	33
8.2.2.2. Merohedral twinning	35
8.2.2.3. Intensity statistics for twined crystals	39
8.2.2.4. Testing for twinning	42
8.3. Fundamentals of data collection	43
8.3.1. Anatomy of a rotation image	45
8.3.2. Data collection as a multi-level decision processes	51
8.3.2.1. Initial inspection of the diffraction image	53
8.3.2.2. Indexing	59
8.3.2.3. Real crystals: modulated structures	63
8.3.2.4. Strategy generation	65
8.3.2.5. Data collection strategy for molecular replacement phasing	65
8.3.2.6. Data collection for high resolution structures	66
8.3.2.7. Anomalous diffraction data	66
8.3.2.8. Single wavelength anomalous data	67
8.3.2.9. Multiple anomalous diffraction data	68
8.3.2.10. Raw data correction, integration and scaling	69
8.3.3. Data reduction and quality analysis	70
8.3.3.1. Merging statistics and redundancy	71
8.3.3.2. Signal-to-noise ratio, completeness and merging <i>R</i> -value	74
8.3.3.3. Space group determination	77

8.3.3.4. Reindexing of unit cells.....	78
8.4. Processing of S-SAD data - a practical example	80
8.4.1. High resolution synchrotron data	81
8.4.1.1. Indexing and strategy generation.....	82
8.4.1.2. Cell refinement, integration and scaling	85
8.4.1.3. Laue symmetry and space group determination	87
8.4.1.4. Reprocessing in correct Laue symmetry	89
8.4.2. Collecting highly redundant anomalous data	89
8.4.2.1. Strategy for highly redundant data collection	91
8.4.2.2. Scaling and merging statistics.....	93
8.5. First analysis and descriptive statistics of data sets	94
8.5.1. Wilson plots and checks for anisotropy.....	94
8.5.2. Testing for twinning.....	95
8.5.3. Determining the unit cell contents.....	96
8.5.4. Analysis of non-crystallographic symmetry.....	97
8.5.5. Assembling 'Table 1'	98
8.6. Next steps	100
8.7. Key concepts	100
Instrumentation	100
Cryo-techniques and real crystals.....	101
Data collection	102
Data quality and analysis	103
8.8. Thought questions	104
8.9. Additional reading.....	104
8.10. References	105

Part IV: Determining your structure

Chapter 9. Reconstruction of electron density and the phase problem

- A basic understanding of how maps are generated and of the phase problem is needed before we can start to learn how to generate phases and solve the structure.

9. Reconstruction of electron density and the phase problem.....	1
9.1. The reconstruction of electron density	3
9.1.1. Fourier transformations.....	4
9.1.2. Computing the Fourier integrals and summations.....	5
9.1.3. The Fourier convolution theorem.....	9
9.1.4. Calculation of electron density	11
9.1.5. Fourier transforms, truncations, and grids.....	13
9.1.6. Effect of atom occupation and <i>B</i> -factor on electron density	17
9.1.7. Three-dimensional electron density maps	18
9.1.8. Resolution and information content of electron density maps	20

9.2. Effects of incomplete data on electron density reconstruction	23
9.2.1. Randomly missing reflections	23
9.2.2. Missing strong reflections	25
9.2.3. Missing data wedges	26
9.2.4. Manifestation of lattice disorder and anisotropy	27
9.3. The phase problem and phases in electron reconstruction	30
9.3.1. Effect of intensity errors on electron density reconstruction	31
9.3.2. Effect of phase errors on electron density reconstruction	32
9.3.3. Phase- or model-bias	35
9.3.4. Basic electron density map types and map coefficients	38
9.3.5. Difference maps	39
9.4. Patterson function and Patterson maps	42
9.4.1. Auto-correlation of the electron density	43
9.4.2. Centrosymmetry of Patterson space	46
9.4.3. Number and width of Patterson peaks	46
9.4.4. Applications of the Patterson function	47
9.5. Solving the phase problem - a preview	48
9.6. Review of key concepts	49
Fourier transforms and maps	49
Phase problem and map reconstruction	50
Patterson function and Patterson maps	50
9.7. Thought questions	51
9.8. Additional Reading	51
9.9. References	51

Chapter 10. Experimental phasing

- Begins gently with pictures and phasing circles and progresses into full maximum likelihood treatment. A quick example at the end reinforces the learning objectives by application.

10.	Experimental phasing	1
10.1.	Solving the phase problem	4
10.2.	Marker atom substructure methods and electronic differences.....	5
10.2.1.	Isomorphous differences.....	8
10.2.2.	Anomalous and dispersive differences	8
10.2.3.	Non-isomorphism and estimate of isomorphous signal change	9
10.2.4.	Obtaining initial protein phases	10
10.3.	Marker atom substructure solution	13
10.3.1.	Preparing the difference data	15
10.3.1.1.	Local scaling	15
10.3.1.2.	Anisotropic scaling.....	16
10.3.1.3.	Outlier detection and resolution truncation	17
10.3.2.	Difference Patterson maps.....	18
10.3.3.	Harker sections and distance vectors.....	21
10.3.4.	Manual solution of a difference Patterson map	21
10.3.4.1.	Location of Harker sections and Patterson vectors	22
10.3.4.2.	Multiple origins and Cheshire groups	25
10.3.4.3.	Heavy atom cross-Fourier maps.....	26
10.3.5.	Automated Patterson search methods	28
10.3.5.1.	Real space grid searches.....	28
10.3.5.2.	Patterson vector methods.....	29
10.3.5.3.	Reciprocal space searches, Fourier methods, combined techniques	30
10.3.6.	Marker atom substructure solution by Direct Methods	31
10.3.6.1.	Ab-initio structure determination versus substructure solution	31
10.3.6.2.	Triplet relations and tangent formula	32
10.3.6.3.	Dual-space Direct Methods	34
10.4.	Solving the phasing equations.....	39
10.4.1.	SIR phasing and SIR maps	39
10.4.2.	SAD phasing and SAD maps.....	41
10.4.3.	Anomalous signal resolves phase ambiguity.....	43
10.4.4.	Density modification resolves phase ambiguity in SIR/SAD maps	44
10.4.5.	The heavy atom enantiomorph problem	45
10.4.6.	Solving phase ambiguity by MIR	48
10.4.7.	MAD phasing equations	50
10.5.	Basic probability distributions for protein phases	52
10.5.1.	Phase combination and best phases	53

10.5.2. Constructing the initial protein electron density map	56
10.6. Heavy atom refinement and phasing	57
10.6.1. Classical heavy atom refinement	58
10.6.2. Maximum likelihood heavy atom refinement.....	59
10.6.2.1. Assumptions in early heavy atom refinement and phasing	60
10.6.2.2. Accounting for incomplete model knowledge	62
10.6.2.3. Likelihood functions and phase probabilities.....	64
10.6.2.4. The MIR likelihood function and non-isomorphism.....	65
10.6.2.5. SAD, SIRAS and MAD likelihood functions	72
10.6.3. Combining phases from different sources	74
10.7. Improving the initial electron density map	74
10.7.1. Solvent modification and phase extension	75
10.7.2. Solvent flattening and solvent masks.....	76
10.7.3. Solvent Flipping	78
10.7.4. Sphere of influence method.....	79
10.7.5. Histogram matching	80
10.7.6. Maximum likelihood reciprocal space methods	81
10.7.7. Density averaging, dummy atom placing and Free Lunch	82
10.8. Fast track from data to high quality protein map	83
10.8.1. S-SAD phasing with highly redundant in-house data	85
10.8.1.1. Data analysis and preparation - SHELXC	85
10.8.1.2. Searching for heavy atom locations - SHELXD.....	86
10.8.1.3. Phase determination and phase extension - SHELXE	88
10.8.1.4. Submission to model building program	93
10.8.2. MAD phasing with Se substructure.....	94
10.8.2.1. Data preparation and substructure solution.....	95
10.8.2.2. Resolving handedness ambiguity and phase extension.....	96
10.9. The next steps	98
10.10. Review of key concepts	98
Phasing methods	98
Substructure methods.....	98
Heavy atom substructure solution	99
Phasing equations and phase probabilities.....	100
Heavy atom refinement and phasing	101
Density modification techniques.....	101
10.11. Exercises	102
10.12. General references	102
10.13. Specific references	103

Chapter 11. Non-crystallographic symmetry and Molecular replacement

- NCS and molecular replacement and MR are tightly connected, and this is the best point to review general transformations and NCS averaging. The MR chapter is also technically challenging, but a number of examples illustrate the process.

11.	Non-crystallographic symmetry and Molecular Replacement.....	1
11.1.	Non-crystallographic symmetry	3
11.1.1.	Determining the unit cell content	4
11.1.1.1.	Matthews coefficient	4
11.1.1.2.	Matthews probabilities	5
11.1.2.	NCS transformations	8
11.1.2.1.	NCS operators.....	9
11.1.2.2.	Rotations in a Cartesian system	10
11.1.2.3.	Eigenvalues and eigenvector of the direction cosine matrix	14
11.1.2.4.	Polar coordinates.....	15
11.1.2.5.	Crystallographic transformations in Cartesian space.....	17
11.1.2.6.	Types of NCS transformations	20
11.1.2.7.	Translational NCS in real space	21
11.1.2.8.	Manifestation of translational NCS in reciprocal space	22
11.1.2.9.	NCS axes parallel to crystallographic axes and native Patterson maps	24
11.1.2.10.	Proper NCS axes and Patterson self-rotation function	26
11.1.2.11.	Finding the NCS operator	28
11.2.	Phase improvement by NCS density averaging.....	29
11.2.1.	Averaging in real space and in reciprocal space	30
11.2.2.	The reciprocal space interference function G	33
11.2.3.	Examples of electron density map averaging	34
11.2.3.1.	Basic map averaging	34
11.2.3.2.	Combined map averaging and density modification	36
11.3.	Molecular Replacement phasing	37
11.3.1.	Placing a real space probe using reciprocal space data	39
11.3.2.	Molecular Replacement by multi-dimensional search	40
11.3.2.1.	Evolutionary search algorithms.....	42
11.3.2.2.	Molecular Replacement of a dimer structure with EPMR	44
11.3.3.	Rotation-translation methods.....	50
11.3.3.1.	Searches in Patterson vector space.....	50
11.3.3.2.	Patterson rotation searches	50
11.3.3.3.	Rotation functions	52
11.3.3.4.	Real space rotation functions.....	53
11.3.3.5.	The fast rotation function	53
11.3.3.6.	Direct rotation function.....	54
11.3.3.7.	Locked rotation functions.....	55
11.3.3.8.	Patterson translation functions.....	56
11.3.3.9.	Translational search space	57

11.3.3.10.	Fast translation functions	57
11.3.3.11.	Phased and locked translation functions, overlap penalties	58
11.3.3.12.	Scoring multiple rotation solutions by translation search	59
11.3.3.13.	Improvement of placement by rigid body refinement	59
11.3.3.14.	Improved translation searches by PC refinement of orientations	60
11.3.4.	Probe selection and search strategies	60
11.3.4.1.	NMR and homology models as MR search models	61
11.3.4.2.	Multiple molecules in the asymmetric unit	61
11.3.4.3.	Automated multi-solution and multi-model searches	62
11.3.5.	Bulk solvent corrections for model structure factors	64
11.3.6.	Maximum likelihood Molecular Replacement	69
11.3.6.1.	A brief review of maximum likelihood functions	69
11.3.6.2.	Translation likelihood functions	70
11.3.6.3.	Likelihood enhanced translation functions	72
11.3.6.4.	Anisotropy correction	72
11.3.6.5.	Maximum likelihood rotation functions	72
11.3.6.6.	Variances, ensembling and normal mode analysis	74
11.3.6.7.	Packing function	75
11.3.6.8.	Automated MR strategies with PHASER	75
11.4.	The next steps	76
11.5.	Review of key concepts	76
	Unit cell contents	76
	Non-crystallographic symmetry operations	76
	NCS in reciprocal space and NCS averaging	77
	Molecular Replacement	77
11.6.	Exercises	79
11.7.	General references	79
11.8.	Specific references	80

Chapter 12 Model building and refinement

- Model building and refinement are at the core of crystallography. Theoretical background including maximum likelihood based techniques and examples of good practices in model building are discussed.

12.	Model building and refinement	1
12.1.	Basics of model building and refinement	4
12.1.1.	The model building and refinement process - an overview	8
12.1.1.1.	Local real space refinement	8
12.1.1.2.	Global reciprocal space restrained refinement	9
12.1.2.	How good or bad are initial phases?	11
12.1.2.1.	Experimental electron density maps	11
12.1.2.2.	Molecular Replacement phases	12

12.1.3. Model errors and phase bias in Molecular Replacement.....	12
12.1.4. Likelihood-based error models	15
12.1.4.1. Review of likelihood functions.....	16
12.1.4.2. Positional errors and incompleteness in σ_A	16
12.1.4.3. Estimating σ_A and figure of merit from experimental data	18
12.1.4.4. Estimating coordinate error and model completeness from σ_A	19
12.1.4.5. Generating bias minimized maximum likelihood maps	19
12.1.4.6. Maximum likelihood map coefficients	20
12.2. Refinement of macromolecular structure models	22
12.2.1. Observations, models, target functions, and algorithms	23
12.2.2. The data to parameter ratio	26
12.2.2.1. Model parameterization	27
12.2.2.2. Crossvalidation and free R-value	28
12.2.2.3. Real space crossvalidation	31
12.2.2.4. Number of observations from experimental diffraction data.....	33
12.2.3. Stereochemical constraints and restraints	35
12.2.3.1. Stereochemical constraints	35
12.2.3.2. Bond distance, bond angle, and dihedral restraints.....	36
12.2.3.3. Planarity and chirality restraints.....	38
12.2.3.4. Side chain versus backbone torsion angle restraints	39
12.2.3.5. Non-bonded interaction restraints.....	40
12.2.3.6. Hydrogen atoms and riding restraints.....	41
12.2.3.7. Non-crystallographic symmetry restraints	42
12.2.3.8. Experimental phase restraints.....	44
12.2.3.9. Displacement parameter (<i>B</i> -factor) restraints	45
12.2.3.10. Restraint setup and data to parameter ratio.....	47
12.2.3.11. Restraint weights and geometry r.m.s.d.	49
12.2.4. Anisotropic displacement and TLS parameterization.....	52
12.2.4.1. Anisotropic diffraction and anisotropic scaling	53
12.2.4.2. TLS description of correlated anisotropic motion	56
12.2.4.3. Implementation of restraints in least squares target functions.....	59
12.2.5. Implementation of restraints in maximum likelihood target functions.....	59
12.2.5.1. Maximum likelihood and stereochemical restraints	60
12.2.5.2. Prior probability function for stereochemical restraints.....	60
12.2.5.3. Likelihood function for the X-ray term	61
12.2.5.4. Model error estimate and likelihood map coefficients	62
12.2.5.5. Incorporating experimental phase restraints in likelihood functions.....	63
12.2.6. Non-stochastic optimization algorithms.....	64
12.2.6.1. Optimization as a general maximum likelihood procedure	64
12.2.6.2. Gradient descent methods	65
12.2.7. Molecular Dynamics, torsion angle refinement, and stochastic optimization.....	68
12.2.7.1. Simulated annealing	69
12.2.7.2. Monte Carlo and molecular dynamics minimization.....	70
12.2.7.3. Simulated annealing torsion angle refinement	71

12.2.8. Dummy atom refinement protocols	72
12.2.8.1. Automated water building.....	73
12.2.8.2. Bias minimized map generation	73
12.2.9. How accurate are protein structures?.....	75
12.2.9.1. Model coordinate precision: the diffraction precision index	76
12.2.9.2. Expectation values for <i>R</i> -free and <i>R</i> -free ratio	80
12.2.10. Summary of refinement programs and protocols	82
12.3. The practice of model building and refinement	85
12.3.1. The basics - manual rebuilding of a Molecular Replacement model	86
12.3.1.1. Correcting and refining the initial MR model	87
12.3.1.2. Building ordered solvent - the hydrogen bond network	90
12.3.1.3. Non-water solvent molecules	92
12.3.1.4. Ligand density	94
12.3.1.5. Ambiguities in side chain electron density - NQH flips.....	95
12.3.1.6. Side chain torsions and multiple side chain conformations.....	96
12.3.1.7. Tracking progress through real space correlation plots	98
12.3.2. Automated model building and improvement	100
12.3.2.1. Automated model rebuilding with <i>ARP/wARP</i>	100
12.3.2.2. Model building into experimental maps	101
12.3.2.3. Electron density of α -helices and β -sheets	104
12.3.2.4. Automated <i>de novo</i> model building	106
12.3.3. Auto-building, correcting and refining a MAD structure.....	108
12.3.4. When is model building and refinement finished?	112
12.3.5. Assembling 'Table 2'	113
12.4. Next steps	115
12.5. Review of key concepts.....	116
12.5.1. Basics of model building and refinement	116
12.5.2. Refinement	117
12.5.3. Model building and refinement practice	119
12.6. Exercises.....	120
12.7. General references	120
12.8. Specific references.....	120

Part V: Making sense of your structure

Chapter 13 Validation, analysis and presentation

- Any crystallographic model presents a molecular hypothesis and thus must be critically evaluated against the body of all prior knowledge. Likelihood based validation provides reliable safeguards against misinterpretation.

13. Structure validation, analysis, and presentation	1
13.1. Beyond mere geometry checking	3
13.1.1. Validation as hypothesis testing	4
13.1.2. Model validation as a responsibility	5
13.1.3. Global versus local structure quality indicators	7
13.2. Examining model stereochemistry	9
13.2.1. <i>PROCHECK</i> stereochemistry analysis	10
13.2.2. <i>MolProbity</i> and all-atom contact analysis	11
13.2.3. Multiple side chain electron density and NQH flips	13
13.2.4. Backbone torsion angle validation	15
13.3. Electron density based model validation.....	17
13.3.1. Real space <i>R</i> -value and real space correlation coefficient	18
13.3.2. The <i>EDS</i> server and <i>SFCHECK</i>	19
13.3.3. Unexpected stereochemistry and real space validation	20
13.4. Validation of protein-ligand complex structures.....	21
13.4.1. Ligand density validation	21
13.4.2. Automated ligand building with <i>ARP/wARP</i>	23
13.4.3. Occupancy and <i>B</i> -factor adjustment	23
13.4.4. Geometry validation and restraint setup for ligands	24
13.4.5. Chemical plausibility of fitted ligands and binding pocket analysis	28
13.4.6. Binding pocket prediction and virtual ligand screening	32
13.5. Low resolution structures	35
13.6. Case study I: an highly improbable proposition	40
13.7. Case study II: perplexing features abound	44
13.8. Case study III: B-factor distribution plots	46
13.9. Depositing the finished structure	47
13.9.1. Deposition checklist.....	48
13.9.2. Deposition pre-check	49
13.9.3. Full validation report.....	51
13.10. Publication and presentation.....	51
13.10.1. Marketing.....	51
13.10.2. Title and abstract	52
13.10.3. The art of scientific illustration	52
13.10.4. Surviving and profiting from review	54
13.11. Review of key concepts.....	57
13.12. Additional reading.....	58
13.13. References	59

Appendix

Contains various ancillary materials the exceed side bar volume or tabular material
Math appendices)

Appendix	1
A.1. Coordinate file formats	1
A.1.1. The PDB file header section.....	2
A.1.2. The CRYST1 and SCALE records	2
A.1.3. The coordinate section of the PDB file	3
A.2. Basic linear algebra	5
A.1.1. Matrix representations.....	6
A.1.2. Sum, difference, and multiplication of matrices	6
A.1.3. Transpose, inversion, determinant, and trace of a matrix.....	7
A.1.4. Vector operations as matrix operations.....	8
A.3. Crystallographic coordinate system and fractional coordinates	10
A.1.1. Direction cosines	11
A.1.2. The metric tensor G, bond distances and bond angles.....	11
A.1.3. Basis transformations	13
A.1.4. Reindexing of diffraction data	13
A.1.5. Coordinate system transformations.....	14
A.1.6. Reciprocal space metric relations.....	15
A.1.7. Reciprocal lattice vectors	16
A.4. Cosine rule.....	18
A.5. Proof of Euler's formula	19
A.6. Electron density is real	20
A.7. Directional cosine matrix in spherical coordinates.....	21
A.8. Crystallographic restriction theorem	21
A.9. Anatomy of a ligand restraint file	22
A.10. The 44 indexing possibilities and possible space groups	24
A.11. Protein – metal coordination and distances	25
A.12. Selected physical constants and conversions	26
A.13. References	26

Glossary**Index**