

Phasing in Crystallography

A Modern Perspective

CARMELO GIACOVAZZO

Professor of Crystallography,

University of Bari, Italy

Institute of Crystallography, CNR, Bari, Italy

OXFORD
UNIVERSITY PRESS

Contents

Symbols and notation

xvii

1	Fundamentals of crystallography	1
1.1	Introduction	1
1.2	Crystals and crystallographic symmetry in direct space	1
1.3	The reciprocal space	5
1.4	The structure factor	11
1.5	Symmetry in reciprocal space	12
1.5.1	Friedel law	12
1.5.2	Effects of symmetry operators in reciprocal space	12
1.5.3	Determination of reflections with restricted phase values	13
1.5.4	Systematic absences	15
1.6	The basic postulate of structural crystallography	17
1.7	The legacy of crystallography	24
2	Wilson statistics	27
2.1	Introduction	27
2.2	Statistics of the structure factor: general considerations	28
2.3	Structure factor statistics in $P1$ and $P\bar{1}$	29
2.4	The $P(z)$ distributions	35
2.5	Cumulative distributions	35
2.6	Space group identification	36
2.7	The centric or acentric nature of crystals: Wilson statistical analysis	42
2.8	Absolute scaling of intensities: the Wilson plot	43
2.9	Shape of the Wilson plot	47
2.10	Unit cell content	49
	Appendix 2.A Statistical calculations in $P1$ and $P\bar{1}$	50
	2.A.1 Structure factor statistics in $P1$	50
	2.A.2 Structure factor statistics in $P\bar{1}$	52
	Appendix 2.B Statistical calculations in any space group	53
	2.B.1 The algebraic form of the structure factor	53
	2.B.2 Structure factor statistics for centric and acentric space groups	55
	Appendix 2.C The Debye formula	58
3	The origin problem, invariants, and seminvariants	60
3.1	Introduction	60
3.2	Origin, phases, and symmetry operators	61

3.3	The concept of structure invariant	63
3.4	Allowed or permissible origins in primitive space groups	65
3.5	The concept of structure seminvariant	69
3.6	Allowed or permissible origins in centred cells	76
3.7	Origin definition by phase assignment	81
4	The method of joint probability distribution functions, neighbourhoods, and representations	83
4.1	Introduction	83
4.2	Neighbourhoods and representations	87
4.3	Representations of structure seminvariants	89
4.4	Representation theory for structure invariants extended to isomorphous data	91
	Appendix 4.A The method of structure factor joint probability distribution functions	93
4.A.1	Introduction	93
4.A.2	Multivariate distributions in centrosymmetric structures: the case of independent random variables	94
4.A.3	Multivariate distributions in non-centrosymmetric structures: the case of independent random variables	97
4.A.4	Simplified joint probability density functions in the absence of prior information	99
4.A.5	The joint probability density function when some prior information is available	102
4.A.6	The calculation of $P(E)$ in the absence of prior information	103
5	The probabilistic estimation of triplet and quartet invariants	104
5.1	Introduction	104
5.2	Estimation of the triplet structure invariant via its first representation: the $P1$ and the $P\bar{1}$ case	104
5.3	About triplet invariant reliability	108
5.4	The estimation of triplet phases via their second representation	110
5.5	Introduction to quartets	112
5.6	The estimation of quartet invariants in $P1$ and $P\bar{1}$ via their first representation: Hauptman approach	112
5.7	The estimation of quartet invariants in $P1$ and $P\bar{1}$ via their first representation: Giacovazzo approach	115
5.8	About quartet reliability	116
	Appendix 5.A The probabilistic estimation of the triplet invariants in $P1$	117
	Appendix 5.B Symmetry inconsistent triplets	120
	Appendix 5.C The P_{10} formula	121
	Appendix 5.D The use of symmetry in quartet estimation	123

6	Traditional direct phasing procedures	125
6.1	Introduction	125
6.2	The tangent formula	128
6.3	Procedure for phase determination via traditional direct methods	130
6.3.1	Set-up of phase relationships	131
6.3.2	Assignment of starting phases	134
6.3.3	Phase determination	136
6.3.4	Finding the correct solution	137
6.3.5	E-map interpretation	138
6.3.6	Phase extension and refinement: reciprocal space techniques	140
6.3.7	The limits of the tangent formula	141
6.4	Third generation direct methods programs	144
6.4.1	The shake and bake approach	144
6.4.2	The half-bake approach	147
6.4.3	The SIR2000-N approach	148
	Appendix 6.A Finding quartets	149
7	Joint probability distribution functions when a model is available: Fourier syntheses	151
7.1	Introduction	151
7.2	Estimation of the two-phase structure invariant ($\phi_h - \phi_{ph}$)	152
7.3	Electron density maps	155
7.3.1	The ideal Fourier synthesis and its properties	156
7.3.2	The observed Fourier synthesis	162
7.3.3	The difference Fourier synthesis	164
7.3.4	Hybrid Fourier syntheses	166
7.4	Variance and covariance for electron density maps	168
7.5	Triplet phase estimate when a model is available	170
	Appendix 7.A Estimation of σ_A	173
	Appendix 7.B Variance and covariance expressions for electron density maps	174
	Appendix 7.C Some marginal and conditional probabilities of $P(R, R_p, \phi, \phi_p)$	176
8	Phase improvement and extension	177
8.1	Introduction	177
8.2	Phase extension and refinement via direct space procedures: EDM techniques	177
8.3	Automatic model building	184
8.4	Applications	188
	Appendix 8.A Solvent content, envelope definition, and solvent modelling	190
8.A.1	Solvent content according to Matthews	190
8.A.2	Envelope definition	191
8.A.3	Models for the bulk solvent	192
	Appendix 8.B Histogram matching	193
	Appendix 8.C A brief outline of the <i>ARP/wARP</i> procedure	196

9	Charge flipping and <i>VLD</i> (<i>vive la difference</i>)	198
9.1	Introduction	198
9.2	The charge flipping algorithm	199
9.3	The <i>VLD</i> phasing method	201
9.3.1	The algorithm	201
9.3.2	<i>VLD</i> and hybrid Fourier syntheses	205
9.3.3	<i>VLD</i> applications to ab initio phasing	205
	Appendix 9.A About <i>VLD</i> joint probability distributions	206
9.A.1	The <i>VLD</i> algorithm based on difference Fourier synthesis	206
9.A.2	The <i>VLD</i> algorithm based on hybrid Fourier syntheses	211
	Appendix 9.B The <i>RELAX</i> algorithm	212
10	Patterson methods and direct space properties	214
10.1	Introduction	214
10.2	The Patterson function	215
10.2.1	Mathematical background	215
10.2.2	About interatomic vectors	216
10.2.3	About Patterson symmetry	217
10.3	Deconvolution of Patterson functions	218
10.3.1	The traditional heavy-atom method	219
10.3.2	Heavy-atom search by translation functions	220
10.3.3	The method of implication transformations	221
10.3.4	Patterson superposition methods	223
10.3.5	The <i>C</i> -map and superposition methods	225
10.4	Applications of Patterson techniques	227
	Appendix 10.A Electron density and phase relationships	230
	Appendix 10.B Patterson features and phase relationships	232
11	Phasing via electron and neutron diffraction data	234
11.1	Introduction	234
11.2	Electron scattering	235
11.3	Electron diffraction amplitudes	236
11.4	Non-kinematical character of electron diffraction amplitudes	237
11.5	A traditional experimental procedure for electron diffraction studies	239
11.6	Electron microscopy, image processing, and phasing methods	241
11.7	New experimental approaches: precession and rotation cameras	244
11.8	Neutron scattering	245
11.9	Violation of the positivity postulate	247
	Appendix 11.A About the elastic scattering of electrons: the kinematical approximation	249
12	Phasing methods for powder data	252
12.1	Introduction	252
12.2	About the diffraction pattern: peak overlapping	253

12.3	Modelling the diffraction pattern	258
12.4	Recovering $ F_{hkl} ^2$ from powder patterns	260
12.5	The amount of information in a powder diagram	263
12.6	Indexing of diffraction patterns	264
12.7	Space group identification	266
12.8	Ab initio phasing methods	267
12.9	Non-ab initio phasing methods	270
	Appendix 12.A Minimizing texture effects	272
13	Molecular replacement	275
13.1	Introduction	275
13.2	About the search model	277
13.3	About the six-dimensional search	279
13.4	The algebraic bases of vector search techniques	280
13.5	Rotation functions	282
13.6	Practical aspects of the rotation function	284
13.7	The translation functions	286
13.8	About stochastic approaches to MR	289
13.9	Combining MR with ‘trivial’ prior information: the <i>ARCIMBOLDO</i> approach	289
13.10	Applications	291
	Appendix 13.A Calculation of the rotation function in orthogonalized crystal axes	294
13.A.1	The orthogonalization matrix	294
13.A.2	Rotation in Cartesian space	295
13.A.3	Conversion to fractional coordinates	297
13.A.4	Symmetry and the rotation function	299
	Appendix 13.B Non-crystallographic symmetry	304
13.B.1	NCS symmetry operators	304
13.B.2	Finding NCS operators	305
13.B.3	The translational NCS	308
	Appendix 13.C Algebraic forms for the rotation and translation functions	311
14	Isomorphous replacement techniques	314
14.1	Introduction	314
14.2	Protein soaking and co-crystallization	315
14.3	The algebraic bases of SIR techniques	317
14.4	The algebraic bases of MIR techniques	320
14.5	Scaling of experimental data	322
14.6	The probabilistic approach for the SIR case	323
14.7	The probabilistic approach for the MIR case	327
14.8	Applications	329
	Appendix 14.A The SIR case for centric reflections	330
	Appendix 14.B The SIR case: the one-step procedure	331
	Appendix 14.C About methods for estimating the scattering power of the heavy-atom substructure	333

15 Anomalous dispersion techniques	335
15.1 Introduction	335
15.2 Violation of the Friedel law as basis of the phasing method	337
15.3 Selection of dispersive atoms and wavelengths	340
15.4 Phasing via SAD techniques: the algebraic approach	344
15.5 The SIRAS algebraic bases	347
15.6 The MAD algebraic bases	352
15.7 The probabilistic approach for the SAD-MAD case	354
15.8 The probabilistic approach for the SIRAS-MIRAS case	360
15.9 Anomalous dispersion and powder crystallography	363
15.10 Applications	364
Appendix 15.A A probabilistic formula for the SAD case	365
Appendix 15.B Structure refinement for MAD data	366
Appendix 15.C About protein phase estimation in the SIRAS case	368
Appendices	370
Appendix M.A Some basic results in probability theory	370
M.A.1 Probability distribution functions	370
M.A.2 Moments of a distribution	371
M.A.3 The characteristic function	371
M.A.4 Cumulants of a distribution	373
M.A.5 The normal or Gaussian distribution	374
M.A.6 The central limit theorem	375
M.A.7 Multivariate distributions	375
M.A.8 Evaluation of the moments in structure factor distributions	377
M.A.9 Joint probability distributions of the signs of the structure factors	379
M.A.10 Some measures of location and dispersion in the statistics of directional data	380
Appendix M.B Moments of the P(Z) distributions	382
Appendix M.C The gamma function	382
Appendix M.D The Hermite and Laguerre polynomials	383
Appendix M.E Some results in the theory of Bessel functions	385
M.E.1 Bessel functions	385
M.E.2 Generalized hypergeometric functions	389
Appendix M.F Some definite integrals and formulas of frequent application	390
<i>References</i>	394
<i>Index</i>	412