

Contents

Contents	i
List of Figures	v
List of Tables	ix
Acknowledgements	xi
Abstract	xiii
1 Introduction	1
1.1 Three Dimensional Model Determination of Macromolecular Assemblies .	1
1.1.1 X-Ray Crystallography and Cryo-Electron Microscopy	2
1.2 Electron Microscopy	3
1.2.1 The Transmission Electron Microscope	4
1.2.2 Image Acquisition	6
1.2.2.1 Amplitude Contrast	7
1.2.2.2 Phase Contrast	8
1.2.2.3 Phase Contrast Transfer Function	11
1.3 Single Particle Electron Microscopy Workflow	13
1.3.1 Laboratory Part of the Workflow	13
1.3.1.1 Sample Treatment	14
1.3.1.2 Grid Preparation	14
1.3.1.2.1 Negative Stain Preparation	15
1.3.1.2.2 Cryo Preparation	15
1.3.1.3 Image Acquisition	16
1.3.1.4 Tilt Images	17
1.3.2 Computational Image Processing	18
1.3.2.1 Particle Picking	18
1.3.2.2 Filtering	19
1.3.2.3 Alignment	20
1.3.2.3.1 Multi Reference Alignment	21
1.3.2.3.2 Reference Free Alignment	22

1.3.2.4	Multivariate Statistical Analysis	23
1.3.2.4.1	Principal Component Analysis	23
1.3.2.4.2	Clustering	25
1.3.2.5	Angular Assignment	27
1.3.2.5.1	Angular Reconstitution	27
1.3.2.5.2	Random Conical Tilt	29
1.3.2.6	3D Reconstruction	29
1.3.2.7	Projection	31
1.3.2.8	Resolution Assessment	32
1.3.2.9	Validation	34
1.3.2.10	Iterative Refinement	35
1.4	Challenges	36
1.4.1	Biochemical Control Over the Sample	36
1.4.2	Model Bias during Refinement	37
1.4.3	Overfitting Induced by Refinement	38
1.4.4	Quality Assessment of a Three-Dimensional Model	39
1.5	Aim of this Work	40
2	Materials and Methods	41
2.1	Fundamentals	41
2.1.1	Image Statistics	41
2.1.1.1	Mean	41
2.1.1.2	Variance	42
2.1.1.3	Standard Deviation	42
2.1.1.4	Signal to Noise Ratio	42
2.1.2	Image Processing Techniques	43
2.1.2.1	Normalization	43
2.1.2.2	Threshold	43
2.1.2.3	Binarization	44
2.1.2.4	Scaling	44
2.1.2.5	Fourier Transformation	45
2.1.2.6	Convolution	45
2.1.2.7	Cross-Correlation	46
2.2	Software	47
2.2.1	Libraries	47
2.2.2	Integrated Development Environment	47
2.2.3	Project Management Tools	48
2.3	Hardware	48
2.4	Chemicals	48

2.5	COW Framework	50
2.5.1	COW Application Programming Interface	50
2.5.2	COW Graphical User Interface	52
2.5.3	COW Server	53
2.6	Generation of Test Data	54
2.6.1	Real Data	54
2.6.2	Synthetic Data	54
3	Results	57
3.1	Sample Optimization	57
3.1.1	Mathematical Model of ProteoPlex	59
3.1.2	Screening Procedure of ProteoPlex	62
3.1.3	Data Analysis of ProteoPlex	62
3.1.4	ProteoPlex Software	64
3.1.4.1	Software Architecture	66
3.2	Model Validation	68
3.2.1	Workflow	68
3.2.1.1	Prerequisites	69
3.2.1.2	Angular Assignment	69
3.2.1.3	Reconstruction	70
3.2.1.4	3D Comparison	71
3.2.1.5	Data Evaluation	71
3.2.2	Test Data Preparation	72
3.2.3	Results of Model Validation	73
3.2.3.1	Heterogeneous Data	74
3.2.3.2	Preferred View	75
3.3	Overfitting Analysis	78
3.3.1	Workflow	78
3.3.2	Prerequisites	78
3.3.2.1	Nail Mask Creation	80
3.3.2.2	Application of the Nail Mask	81
3.3.2.3	Evaluation of the Nail Mask Content	82
3.3.3	Results of the Overfitting Analysis	82
3.4	Model Quality Criteria	86
3.4.1	Model Quality Criteria in Detail	86
3.4.1.1	Signal to Noise Ratio	86
3.4.1.2	Sphere Coverage	88
3.4.1.3	Continuity	91
3.4.1.4	Isotropy	93

3.4.1.5	Global FSC	96
3.4.1.6	Local FSC	96
3.4.1.7	Oversampling	97
3.4.1.8	Tilted Pair Parameter Plot	97
3.4.1.9	3D Correlation of Tilted Pairs	98
3.4.2	Model Quality Criteria Results	99
4	Discussion	105
4.1	Sample Optimization Discussion	105
4.1.1	Sample Optimization Outlook	107
4.2	Model Validation Discussion	108
4.2.1	Model Validation Outlook	111
4.3	Overfitting Discussion	113
4.3.1	Overfitting Outlook	114
4.4	Model Quality Criteria Discussion	116
4.4.1	Model Quality Criteria Outlook	118
4.5	Conclusion	119
Appendices		
A	Sample Optimization	I
B	Model Validation Plots	III
C	Quality Criteria Sample Table Content	VII
D	Abbreviations	XV
Bibliography		XIX
Curriculum Vitae		XXVII
Related Publications		XXIX