NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.
ANALYSIS AND PROCESSING
OF DARK-FIELD ELECTRON MICROGRAPHS FOR
THREE DIMENSIONAL RECONSTRUCTION

by

Neil Alexander Farrow

A thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy,
Graduate Department of Medical Biophysics, in the
University of Toronto

© Copyright by Neil Alexander Farrow 1992
The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

ISBN 0-315-78720-1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**THE HUMANITIES AND SOCIAL SCIENCES**

<table>
<thead>
<tr>
<th>Subject Categories</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMUNICATIONS AND THE ARTS</td>
<td>0779</td>
</tr>
<tr>
<td>Architecture</td>
<td>0777</td>
</tr>
<tr>
<td>Art History</td>
<td>0737</td>
</tr>
<tr>
<td>Cinema</td>
<td>0900</td>
</tr>
<tr>
<td>Dance</td>
<td>0905</td>
</tr>
<tr>
<td>Fine Arts</td>
<td>0357</td>
</tr>
<tr>
<td>Information Sciences</td>
<td>0353</td>
</tr>
<tr>
<td>Journalism</td>
<td>0391</td>
</tr>
<tr>
<td>Library Science</td>
<td>0399</td>
</tr>
<tr>
<td>Mass Communications</td>
<td>0708</td>
</tr>
<tr>
<td>Music</td>
<td>0413</td>
</tr>
<tr>
<td>Speech Communication</td>
<td>0459</td>
</tr>
<tr>
<td>Theater</td>
<td>0463</td>
</tr>
<tr>
<td>EDUCATION</td>
<td>0515</td>
</tr>
<tr>
<td>General</td>
<td>0514</td>
</tr>
<tr>
<td>Administration</td>
<td>0515</td>
</tr>
<tr>
<td>Adult and Continuing</td>
<td>0516</td>
</tr>
<tr>
<td>Agronomy and Natural History of Agriculture</td>
<td>0517</td>
</tr>
<tr>
<td>Art</td>
<td>0273</td>
</tr>
<tr>
<td>Bilingual and Multicultural</td>
<td>0275</td>
</tr>
<tr>
<td>Business</td>
<td>0785</td>
</tr>
<tr>
<td>Community College</td>
<td>0795</td>
</tr>
<tr>
<td>Curriculum and Instruction</td>
<td>0777</td>
</tr>
<tr>
<td>Early Childhood Education</td>
<td>0318</td>
</tr>
<tr>
<td>Elementary</td>
<td>0524</td>
</tr>
<tr>
<td>French</td>
<td>0519</td>
</tr>
<tr>
<td>Guidance and Counseling</td>
<td>0525</td>
</tr>
<tr>
<td>Health</td>
<td>0506</td>
</tr>
<tr>
<td>Higher Education</td>
<td>0510</td>
</tr>
<tr>
<td>History of Education</td>
<td>0520</td>
</tr>
<tr>
<td>Home Economics</td>
<td>0278</td>
</tr>
<tr>
<td>Industrial Education</td>
<td>0521</td>
</tr>
<tr>
<td>Language and Literature</td>
<td>0279</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0280</td>
</tr>
<tr>
<td>Music</td>
<td>0522</td>
</tr>
<tr>
<td>Philosophy</td>
<td>0998</td>
</tr>
<tr>
<td>Physical</td>
<td>0523</td>
</tr>
</tbody>
</table>

**LINGUISTICS**

<table>
<thead>
<tr>
<th>Language</th>
<th>0524</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>0519</td>
</tr>
<tr>
<td>Ancient</td>
<td>0279</td>
</tr>
<tr>
<td>Classical</td>
<td>0290</td>
</tr>
<tr>
<td>Comparative</td>
<td>0294</td>
</tr>
<tr>
<td>Medieval</td>
<td>0506</td>
</tr>
<tr>
<td>Modern</td>
<td>0529</td>
</tr>
<tr>
<td>African</td>
<td>0316</td>
</tr>
<tr>
<td>American</td>
<td>0503</td>
</tr>
<tr>
<td>Canadian</td>
<td>0312</td>
</tr>
<tr>
<td>Germanic</td>
<td>0511</td>
</tr>
<tr>
<td>Latin American</td>
<td>0313</td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>0312</td>
</tr>
<tr>
<td>Romance</td>
<td>0314</td>
</tr>
<tr>
<td>Slavic and East European</td>
<td>0313</td>
</tr>
</tbody>
</table>

**PHILOSOPHY, RELIGION AND THEOLOGY**

<table>
<thead>
<tr>
<th>Philosophy</th>
<th>0422</th>
</tr>
</thead>
<tbody>
<tr>
<td>Religion</td>
<td>0525</td>
</tr>
<tr>
<td>Biblical Studies</td>
<td>0321</td>
</tr>
<tr>
<td>Clergy</td>
<td>0819</td>
</tr>
<tr>
<td>History of Philosophy</td>
<td>0320</td>
</tr>
<tr>
<td>Theology</td>
<td>0469</td>
</tr>
</tbody>
</table>

**SOCIAL SCIENCES**

<table>
<thead>
<tr>
<th>American Studies</th>
<th>0323</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthropology</td>
<td>0324</td>
</tr>
<tr>
<td>Archaeology</td>
<td>0326</td>
</tr>
<tr>
<td>Business Administration</td>
<td>0310</td>
</tr>
<tr>
<td>Accounting</td>
<td>0272</td>
</tr>
<tr>
<td>Banking</td>
<td>0770</td>
</tr>
<tr>
<td>Management</td>
<td>0454</td>
</tr>
<tr>
<td>Marketing</td>
<td>0338</td>
</tr>
<tr>
<td>Canadian Studies</td>
<td>0385</td>
</tr>
<tr>
<td>Economics</td>
<td>0501</td>
</tr>
<tr>
<td>Agricultural Economics</td>
<td>0502</td>
</tr>
<tr>
<td>Commerce</td>
<td>0505</td>
</tr>
<tr>
<td>Finance</td>
<td>0508</td>
</tr>
<tr>
<td>History</td>
<td>0509</td>
</tr>
<tr>
<td>History of Science</td>
<td>0585</td>
</tr>
<tr>
<td>History of Technology</td>
<td>0586</td>
</tr>
<tr>
<td>History of Industrial Relations</td>
<td>0629</td>
</tr>
<tr>
<td>Public and Social Welfare</td>
<td>0630</td>
</tr>
<tr>
<td>Social Structure and Development</td>
<td>0700</td>
</tr>
<tr>
<td>Urban and Regional Planning</td>
<td>0709</td>
</tr>
<tr>
<td>Women's Studies</td>
<td>0453</td>
</tr>
</tbody>
</table>

**THE SCIENCES AND ENGINEERING**

<table>
<thead>
<tr>
<th>Subject Categories</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOLOGICAL SCIENCES</td>
<td>0473</td>
</tr>
<tr>
<td>General</td>
<td>0568</td>
</tr>
<tr>
<td>Agronomy</td>
<td>0285</td>
</tr>
<tr>
<td>Animal Culture and Management</td>
<td>0476</td>
</tr>
<tr>
<td>Animal Pathology</td>
<td>0359</td>
</tr>
<tr>
<td>Food and Food Technology</td>
<td>0561</td>
</tr>
<tr>
<td>Forestry and Wildlife</td>
<td>0487</td>
</tr>
<tr>
<td>Plant Pathology</td>
<td>0480</td>
</tr>
<tr>
<td>Plant Physiology</td>
<td>0617</td>
</tr>
<tr>
<td>Range Management</td>
<td>0702</td>
</tr>
<tr>
<td>Wood Technology</td>
<td>0746</td>
</tr>
<tr>
<td>Biology</td>
<td>0300</td>
</tr>
<tr>
<td>Anatomy</td>
<td>0287</td>
</tr>
<tr>
<td>Biostatistics</td>
<td>0350</td>
</tr>
<tr>
<td>Botany</td>
<td>0519</td>
</tr>
<tr>
<td>Cell</td>
<td>0379</td>
</tr>
<tr>
<td>Ecology</td>
<td>0279</td>
</tr>
<tr>
<td>Entomology</td>
<td>0369</td>
</tr>
<tr>
<td>Genetics</td>
<td>0793</td>
</tr>
<tr>
<td>Limnology</td>
<td>0513</td>
</tr>
<tr>
<td>Molecular Biology</td>
<td>0500</td>
</tr>
<tr>
<td>Neurosciences</td>
<td>0506</td>
</tr>
<tr>
<td>Oceanography</td>
<td>0416</td>
</tr>
<tr>
<td>Physiology</td>
<td>0433</td>
</tr>
<tr>
<td>Radiology</td>
<td>0491</td>
</tr>
<tr>
<td>Veterinary Sciences</td>
<td>0708</td>
</tr>
<tr>
<td>Zoology</td>
<td>0472</td>
</tr>
<tr>
<td>Biophysics</td>
<td>0876</td>
</tr>
<tr>
<td>General</td>
<td>0563</td>
</tr>
<tr>
<td>Medical</td>
<td>0670</td>
</tr>
<tr>
<td>EARTH SCIENCES</td>
<td>0423</td>
</tr>
<tr>
<td>Geophysics</td>
<td>0996</td>
</tr>
</tbody>
</table>

**PHYSICAL SCIENCES**

<table>
<thead>
<tr>
<th>Pure Sciences</th>
<th>0485</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>0328</td>
</tr>
<tr>
<td>General</td>
<td>0329</td>
</tr>
<tr>
<td>Analytical Chemistry</td>
<td>0486</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>0487</td>
</tr>
<tr>
<td>Inorganic Chemistry</td>
<td>0488</td>
</tr>
<tr>
<td>Nuclear Chemistry</td>
<td>0738</td>
</tr>
<tr>
<td>Organic Chemistry</td>
<td>0490</td>
</tr>
<tr>
<td>Physical Chemistry</td>
<td>0491</td>
</tr>
<tr>
<td>Polymer Chemistry</td>
<td>0493</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0403</td>
</tr>
<tr>
<td>Physics</td>
<td>0405</td>
</tr>
<tr>
<td>General</td>
<td>0505</td>
</tr>
<tr>
<td>Acoustics</td>
<td>0856</td>
</tr>
<tr>
<td>Astronomy and Astrophysics</td>
<td>0606</td>
</tr>
<tr>
<td>Atmospheric Science</td>
<td>0608</td>
</tr>
<tr>
<td>Atomic Physics</td>
<td>0748</td>
</tr>
<tr>
<td>Electronics and Electrical Engineering</td>
<td>0607</td>
</tr>
<tr>
<td>Helminths</td>
<td>0430</td>
</tr>
<tr>
<td>High Energy Physics</td>
<td>0798</td>
</tr>
<tr>
<td>Fluid and Plasma Physics</td>
<td>0799</td>
</tr>
<tr>
<td>Nuclear Physics</td>
<td>0609</td>
</tr>
<tr>
<td>Physics of the Atom</td>
<td>0610</td>
</tr>
<tr>
<td>Optics</td>
<td>0434</td>
</tr>
<tr>
<td>Radioactivity</td>
<td>0756</td>
</tr>
<tr>
<td>Solid State Physics</td>
<td>0611</td>
</tr>
<tr>
<td>Statistics</td>
<td>0483</td>
</tr>
<tr>
<td>Applied Sciences</td>
<td>0346</td>
</tr>
<tr>
<td>Applied Mechanics</td>
<td>0560</td>
</tr>
<tr>
<td>Computer Science</td>
<td>0984</td>
</tr>
</tbody>
</table>

**ENGINEERING**

<table>
<thead>
<tr>
<th>General</th>
<th>0457</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural Engineering</td>
<td>0338</td>
</tr>
<tr>
<td>Automotive Engineering</td>
<td>0359</td>
</tr>
<tr>
<td>Biomedical Engineering</td>
<td>0540</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td>0541</td>
</tr>
<tr>
<td>Civil Engineering</td>
<td>0352</td>
</tr>
<tr>
<td>Electronics and Electrical Engineering</td>
<td>0544</td>
</tr>
<tr>
<td>Heat and Thermo-fluids Engineering</td>
<td>0348</td>
</tr>
<tr>
<td>Hydraulic Engineering</td>
<td>0349</td>
</tr>
<tr>
<td>Industrial Engineering</td>
<td>0545</td>
</tr>
<tr>
<td>Marine Engineering</td>
<td>0546</td>
</tr>
<tr>
<td>Materials Science</td>
<td>0547</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>0548</td>
</tr>
<tr>
<td>Metallurgical Engineering</td>
<td>0743</td>
</tr>
<tr>
<td>Mining Engineering</td>
<td>0551</td>
</tr>
<tr>
<td>Nuclear Engineering</td>
<td>0552</td>
</tr>
<tr>
<td>Packaging Engineering</td>
<td>0549</td>
</tr>
<tr>
<td>Plastics Engineering</td>
<td>0765</td>
</tr>
<tr>
<td>Sanitary and Municipal Engineering</td>
<td>0554</td>
</tr>
<tr>
<td>System Science</td>
<td>0790</td>
</tr>
<tr>
<td>Geophysics</td>
<td>0428</td>
</tr>
<tr>
<td>Geology</td>
<td>0429</td>
</tr>
<tr>
<td>Operations Research</td>
<td>0766</td>
</tr>
<tr>
<td>Plastic Technology</td>
<td>0795</td>
</tr>
<tr>
<td>Textile Technology</td>
<td>0994</td>
</tr>
</tbody>
</table>

**PSYCHOLOGY**

<table>
<thead>
<tr>
<th>General</th>
<th>0621</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral Science</td>
<td>0384</td>
</tr>
<tr>
<td>Clinical Psychology</td>
<td>0622</td>
</tr>
<tr>
<td>Developmental Psychology</td>
<td>0620</td>
</tr>
<tr>
<td>Experimental Psychology</td>
<td>0623</td>
</tr>
<tr>
<td>Industrial Psychology</td>
<td>0624</td>
</tr>
<tr>
<td>Personality Psychology</td>
<td>0625</td>
</tr>
<tr>
<td>Physiological Psychology</td>
<td>0626</td>
</tr>
<tr>
<td>Psychological Assessment</td>
<td>0632</td>
</tr>
<tr>
<td>Social Psychology</td>
<td>0631</td>
</tr>
<tr>
<td>Social Psychology</td>
<td>0632</td>
</tr>
</tbody>
</table>
NOTE: The AUTHOR will sign in one of the two places indicated. It is the intention of the University that there be NO RESTRICTION on the distribution of the publication of theses save in exceptional cases.

(a) Immediate publication in microform by the National Library is authorized.

    Author's signature ___________________________  Date ________________

- OR -

(b) Publication by the National Library is to be postponed until ___________________ 19
    (normal maximum delay is two years). Meanwhile this thesis may not be consulted in the University Library except with written permission on each occasion from me.

    Author's signature ___________________________  Date ________________

This restriction is authorized for reasons which seem to me, as Chair of the Graduate Department of ____________________________, to be sufficient.

    Signature of Graduate Department Chair ____________________________

    Date ________________

BORROWERS undertake to give proper credit for any use made of the thesis, and to obtain the consent of the author if it is proposed to make extensive quotations, or to reproduce the thesis in whole or in part.

<table>
<thead>
<tr>
<th>Signature of borrower</th>
<th>Address</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1989
ABSTRACT

ANALYSIS AND PROCESSING OF DARK-FIELD ELECTRON MICROGRAPHS FOR THREE DIMENSIONAL RECONSTRUCTION

Neil Alexander Farrow


Department of Medical Biophysics

University of Toronto

Dark-field electron microscopy has the potential of providing high resolution images of biological molecules with a resolution of 5 Å. The two dimensional information from dark-field electron micrographs may be used to generate a three dimensional structure of the molecule if the relative projection directions of the images are known. This thesis describes computational techniques aimed at recovering this three dimensional information from single images of molecules which adopt unknown orientations with respect to the imaging beam.

Dark-field images contain large amounts of noise, primarily a result of structure in the carbon film used to support the specimen. The application of maximum entropy methods, in an attempt to enhance the quality of the original images by removing the noise, is described. It is seen that, to be useful, these methods require a number of other constraints. However, the additional constraints lead to significantly increased computational times which render the algorithm impractical.

An algorithm is described that permits the \textit{a posteriori} determination of the relative projection direction of a large number of molecules. The method uses the central section...
theorem to determine the relative orientations of the images. Quaternion mathematics is used to determine optimum alignments of images. A number of algorithm-specific parameters are developed to indicate the degree of success of the alignment procedure.

Simulated images and dark-field electron micrographs of the Klenow Fragment of DNA Polymerase I are used to test the efficacy of the alignment algorithm. The simulations indicate, on the basis of signal-to-ratio as a measure of image quality, that the algorithm should be able to determine projection directions from dark-field micrographs of single molecules. However, application to electron micrographs provides low quality alignments.

Reasons for the discrepancy between the simulations and experimental situations are discussed. Preliminary experiments indicate that visual selection of high quality images yields improved alignments. It is suggested that this improvement results from selection of particles that have retained their integrity during the imaging procedure. It is concluded that the alignment procedure will probably be successful when applied to higher quality dark-field electron micrographs.
## CONTENTS

```
Contents                                               i
List of Figures                                       v
List of Tables                                        vii
Acknowledgements                                      viii

CHAPTER 1: INTRODUCTION

1.1 The Role of Dark-Field Microscopy in Biological Structure Determination  2
1.1.1 Dark-Field Electron Microscopy                        5
1.1.2 The Imaging Environment and Radiation Damage          7
1.2 Image Restoration and Enhancement                      11
1.2.1 Maximum Entropy Methods                              14
1.2.2 Pattern Recognition Techniques Applied to Image Averaging in Electron Microscopy  16
1.2.2.1 Introduction                                       16
1.2.2.2 Automated Image Alignment                          16
1.2.2.3 Multivariate Statistical Analysis                  18
1.2.2.4 Classification                                     20
1.3 Three-Dimensional Structures from Electron Microscopy  22
1.3.1 Analysis of 3D Structures Through Visual Interpretation of Electron Micrographs  22
1.3.2 Determination of the Three-Dimensional Structure of Symmetrical Objects  24
1.3.3 Three-Dimensional Reconstruction Using Tilt Series    26
1.3.4 The Random-Conical Tilt Method                       28
1.3.5 Reconstruction of Randomly Oriented Asymmetrical Particles  29
1.3.6 Previous Approaches to Automatic Orientation Determination  30
1.4 The Klenow Fragment of DNA Polymerase I                33
1.4.1 Previous Studies                                     33
1.4.2 Selection of the Klenow Fragment for Electron Microscopic Studies  36
1.5 Structure of the Thesis                               38
References                                               41
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER 2: MAXIMUM ENTROPY METHODS AND DARK-FIELD MICROSCOPY IMAGES

2.1 Introduction
2.1.1 The $\chi^2$ Constrained Maximum Entropy Method
2.1.2 The $E^2$ Constrained Maximum Entropy Method
2.2 Simulated Annealing Applied to the Entropy Maximization Problem
2.3 Introduction of a Spatial Constraint
2.4 Inclusion of an Autocovariance Constraint
2.5 Further Considerations
2.5.1 The Application of Simulated Annealing
2.5.2 Other Entropy Expressions
2.6 Conclusions

APPENDICES
2.A.1 The $\chi^2$ Constrained Maximum Entropy Method
2.A.2 The $E^2$ Constrained Maximum Entropy Method

REFERENCES 66

CHAPTER 3: A POSTERIORI DETERMINATION OF RELATIVE PROJECTION DIRECTIONS OF ARBITRARILY ORIENTED MACROMOLECULES

3.1 Introduction
3.1.1 The Central Section Theorem
3.1.2 Previous Work with Asymmetrical Objects
3.1.3 Experimental Constraints
3.2 Coordinate Systems and Nomenclature
3.2.1 Euler Angles and Rotation Matrices
3.2.2 Common Axes
3.3 Orientation of Three Projections in a Single Coordinate System
3.4 Addition of a Fourth Projection
3.5 Extension to a Large Number of Projections
3.6 Measures of Algorithm Performance
3.7 Simulations
3.8 Results and Discussion
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9</td>
<td>Conclusion</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Appendices</td>
<td>95</td>
</tr>
<tr>
<td>3.A.1</td>
<td>Alignment of Three Projections</td>
<td>95</td>
</tr>
<tr>
<td>3.A.2</td>
<td>Determination of Rotation Matrices $T$ by Quaternion Methods</td>
<td>98</td>
</tr>
<tr>
<td>3.A.2.1</td>
<td>Properties of Quaternions</td>
<td>98</td>
</tr>
<tr>
<td>3.A.2.2</td>
<td>Application to the Problem of Determining Matrix $T_z$</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>102</td>
</tr>
</tbody>
</table>

CHAPTER 4: AUTOMATIC ALIGNMENT OF PROJECTION IMAGES UNDER ADVERSE CONDITIONS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>105</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Reference-Free Alignment Techniques</td>
<td>107</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Projection Direction Determination</td>
<td>108</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Assessment of Algorithm Performance</td>
<td>109</td>
</tr>
<tr>
<td>4.2</td>
<td>Experimental Simulations</td>
<td>110</td>
</tr>
<tr>
<td>4.3</td>
<td>Results and Discussion</td>
<td>114</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Definitions</td>
<td>114</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Reference-Free Alignment Results</td>
<td>114</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Classification Results</td>
<td>115</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Projection Direction Determination Results</td>
<td>119</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusions</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Appendices</td>
<td>128</td>
</tr>
<tr>
<td>4.A.1</td>
<td>Definition of Optimum Alignment</td>
<td>128</td>
</tr>
<tr>
<td>4.A.2</td>
<td>Measurement of Algorithm Performance</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>131</td>
</tr>
</tbody>
</table>

CHAPTER 5: THREE-DIMENSIONAL STRUCTURAL STUDIES OF THE KLENOW FRAGMENT OF DNA POLYMERASE I USING DARK-FIELD ELECTRON MICROSCOPY

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>135</td>
</tr>
<tr>
<td>5.2</td>
<td>Methods</td>
<td>138</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Electron Microscopy</td>
<td>138</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

**CHAPTER 1**

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Structure of the Klenow Fragment as Determined by X-ray Crystallography.</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>Similarities Between Dark-Field Electron Micrographs of the Klenow Fragment and the X-ray Crystallography Structure.</td>
<td>37</td>
</tr>
</tbody>
</table>

**CHAPTER 2**

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Simulated Dark-Field Image and Its Maximum Entropy Image.</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>Application of a $\chi^2$ Constrained Maximum Entropy Algorithm to a 1D Data Set.</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>Application of an $E^2$ Constrained Maximum Entropy Algorithm to a 1D Data Set.</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>Simulated Annealing $E^2$ Constrained Algorithm Applied to a 1D Step Function.</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>Autocovariance Function of 1D Data, Noise and Maximum Entropy Solution.</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>The Distribution of $\kappa$ and Entropy for a 1D Data Set.</td>
<td>61</td>
</tr>
</tbody>
</table>

**CHAPTER 3**

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Example of Simulated Projection Images and Corresponding Sinograms.</td>
<td>76</td>
</tr>
<tr>
<td>2</td>
<td>Relationship Between Image Noise and Errors in Common Axis Position Determination.</td>
<td>77</td>
</tr>
<tr>
<td>3</td>
<td>Euler Angle System.</td>
<td>79</td>
</tr>
<tr>
<td>4</td>
<td>Alignment of a Fourth Projection.</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>A Scheme for Determining which Common Axes will be Determined.</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>The Effect of the Number of Iterations on the Alignment Solution.</td>
<td>89</td>
</tr>
<tr>
<td>7</td>
<td>The Effect of Error in Common Axis Position on the Alignment Solution.</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>The Effect of the Number of Check Projections on the Alignment Solution.</td>
<td>92</td>
</tr>
<tr>
<td>A.1</td>
<td>Conventions Used to Define Common Axes Position.</td>
<td>95</td>
</tr>
<tr>
<td>A.2</td>
<td>Alignment of First Three Projections in a Single Coordinate System.</td>
<td>96</td>
</tr>
</tbody>
</table>

**CHAPTER 4**

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Examples of Simulated Projection Images.</td>
<td>113</td>
</tr>
<tr>
<td>Fig. 2</td>
<td>Examples of Misregistrations Arising from the Reference-Free Alignment Procedure.</td>
<td>115</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Fig. 3</td>
<td>Examples of Image Misclassification Arising from Multivariate Statistical Analysis.</td>
<td>118</td>
</tr>
<tr>
<td>Fig. 4</td>
<td>IQAD Determined Projection Directions for the Data Set Sim(i3.1).</td>
<td>120</td>
</tr>
<tr>
<td>Fig. 5</td>
<td>Individual Self-Consistency Errors for the Data Set Sim(i3.1).</td>
<td>122</td>
</tr>
<tr>
<td>Fig. 6</td>
<td>mIQAD Determined Projection Directions for the Data Set Sim(i3.1).</td>
<td>124</td>
</tr>
<tr>
<td>Fig. 7</td>
<td>mIQAD Determined Projection Directions for the Data Set Sim(p3.1).</td>
<td>125</td>
</tr>
</tbody>
</table>

**CHAPTER 5**

<table>
<thead>
<tr>
<th>Fig. 1</th>
<th>Schematic of Procedures Required for 3D Reconstruction.</th>
<th>141</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 2</td>
<td>Grey-Level Histogram for 1028 Images.</td>
<td>143</td>
</tr>
<tr>
<td>Fig. 3</td>
<td>An Example of an Area of Electron Micrograph 858.</td>
<td>146</td>
</tr>
<tr>
<td>Fig. 4</td>
<td>Particle Selection and 2D Alignment of Digitized Images.</td>
<td>147</td>
</tr>
<tr>
<td>Fig. 5</td>
<td>Rotation Angles Assigned by the Reference-Free Alignment Procedure.</td>
<td>148</td>
</tr>
<tr>
<td>Fig. 6</td>
<td>Dendrograms Illustrating the Classification of Image Data Sets.</td>
<td>150</td>
</tr>
<tr>
<td>Fig. 7</td>
<td>Examples of Image Classes as Determined by Principal Component Analysis.</td>
<td>151</td>
</tr>
<tr>
<td>Fig. 8</td>
<td>Examples of Sinograms and a Sinogram Correlation Function from the S1 Data Set.</td>
<td>152</td>
</tr>
<tr>
<td>Fig. 9</td>
<td>Distribution of Sinogram Correlation Function Maxima.</td>
<td>153</td>
</tr>
<tr>
<td>Fig. 10</td>
<td>Projection Directions Determined by the IQAD Procedure.</td>
<td>154</td>
</tr>
<tr>
<td>Fig. 11</td>
<td>Individual Self-Consistency Errors from the IQAD Alignment.</td>
<td>155</td>
</tr>
<tr>
<td>Fig. 12</td>
<td>Projection Directions Determined by the mIQAD Procedure.</td>
<td>157</td>
</tr>
<tr>
<td>Fig. 13</td>
<td>Reconstructed Volumes from the IQAD Procedure.</td>
<td>158</td>
</tr>
<tr>
<td>Fig. 14</td>
<td>Schematic of Sinogram Correlation Function to Explain Features of Fig. 8.</td>
<td>161</td>
</tr>
</tbody>
</table>

**CHAPTER 6**

<table>
<thead>
<tr>
<th>Fig. 1</th>
<th>Relationship Between S.C.E., Euler Angle Error and 3D Correlation Coefficient.</th>
<th>172</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 2</td>
<td>Examples of Reconstructions at Different S.C.E. Values.</td>
<td>173</td>
</tr>
<tr>
<td>Fig. 3</td>
<td>Visually Selected Images Used to Generate Volumes Shown in Fig. 6.</td>
<td>179</td>
</tr>
<tr>
<td>Fig. 4</td>
<td>Distribution of Sinogram Correlation Function Maxima for Images in Fig. 3.</td>
<td>180</td>
</tr>
<tr>
<td>Fig. 5</td>
<td>Projection Directions Assigned by the IQAD Procedure to Images in Fig. 3.</td>
<td>181</td>
</tr>
<tr>
<td>Fig. 6</td>
<td>3D Reconstruction From Images in Fig. 3.</td>
<td>183</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LIST OF TABLES

CHAPTER 4

Table 1  Data Set Description and Reference-Free Alignment Results. 112
Table 2  Classification Results. 116
Table 3  IQAD Projection Direction Determination Results. 119
Table 4  mIQAD Projection Direction Determination Results. 123

CHAPTER 5

Table 1  Details of the Numbers of Images Collected. 139
Table 2  S.C.E. Results from the IQAD and mIQAD Alignments. 156
I would like to acknowledge the contributions of the following people to the work described in this thesis.

Dr Peter Ottensmeyer, for his expert guidance, for his willingness to let me explore new ideas and for his encouragement throughout my Doctoral studies.

Drs Mark Henkelman and Jeremy Carver, for their valuable advice and direction as members of my Student Committee.

Brenda Rutherford and Alex Yu, for introducing me to the practical aspects of dark-field electron microscopy techniques.

Dr George Harauz, for many helpful discussions and much useful advice about the application of image processing techniques.

The National Cancer Institute of Canada, the Medical Research Council of Canada, the Ontario Cancer Research and Treatment Foundation, and the Connaught Foundation for their financial support of these studies.

Finally, I dedicate this thesis to my parents, in appreciation of the support and encouragement they have given me over the years.
CHAPTER 1

INTRODUCTION

Neil A. Farrow
1.1 The Role of Electron Microscopy in Biological Structure Determination

Molecular biology is increasingly concerned with understanding the influence of a single element of a biological system on the remainder of that system. In order for the observations of such functional studies to be understood it is usually necessary to conceptualize the influence of the molecule under study in terms of interactions with other elements. Implicit in many of these interpretations is the assumption that the structure and function of the molecule are intimately linked. Structural biology adds to the understanding of the relationship between structure and function in two ways. Firstly, it can serve to permit the confirmation of structurally based hypotheses that are used to explain the results of functional studies. Secondly, the determination of novel structures will, in turn, suggest new structurally based hypotheses.

Electron microscopy has played a significant role in elucidating the structures that underlie biological processes. The wavelength of the electron gives a potential resolution of 0.01 nm or below. However, as a result of lens imperfections and aberrations in the microscope the best practically obtainable resolutions today are around 0.1 nm. The ultimate limit on attainable resolution however, is usually related to specimens and results from chromatic aberrations and from damage during preparation and imaging procedures. Because of these specimen related resolution problems it is unlikely that electron microscopy of biological specimens will ever attain the 0.1-0.2 nm resolution obtained in X-ray crystallography and nuclear magnetic resonance (NMR) studies.

Given that X-ray crystallography and NMR have a higher resolution potential than electron microscopy, what are the advantages of performing studies of biomolecular structures using electron microscopy? Electron microscopy can be performed using very little material, often nanoliter quantities are sufficient. In contrast X-ray crystallography and NMR often require millimolar quantities to perform their studies. Electron microscopy generally requires no further processing of the sample after it has been purified. X-ray crystallography requires that the molecule be crystallized and often heavy atom derivatives must be prepared to permit isomorphous replacement techniques to be employed. Studies using multidimensional NMR techniques, which permit the solution of proteins of molecular weights ≈ 10 kD, require
labelling of the protein via incorporation of $^{15}\text{N}$ and $^{13}\text{C}$ nuclei. The assignment of resonances present in NMR spectra also requires that the primary sequence of the protein is known.

Electron microscopy studies may be performed at a wide range of resolutions. It is often possible to postulate a low resolution structure almost instantaneously and then go on to refine this structure to higher resolution using image processing techniques at a later date. X-ray crystallography can also solve structures over a range of resolutions but the process of determining even a low resolution structure may take a considerable time. Studies with NMR spectroscopy are all at, or close to, atomic resolution, and intermediate resolution 3D structures exist as only a result of insufficient constraints being available to the programs used to determine the final structure. The time for solution of structures in electron microscopy is typically much shorter than that in the other two modalities. However, because electron microscopy is not as widely applied to 3D structure determination problems as the other techniques, there is often a considerable time investment required in setting up the algorithms required for such studies. Electron microscopy experiments and data processing techniques are a lot more diverse than those in the other two fields.

Electron microscopy permits the determination of structures in a non-crystalline state, removing any concerns about crystal-induced structural alterations. The advent of the solution study techniques in NMR, however, makes the environment of the biomolecule in electron microscopy experiments appear extremely harsh by comparison. The problems associated with the imaging environment in electron microscopy are discussed in section 1.1.2 below. The observation of molecules embedded in a layer of vitreous ice coupled with cryo-electron microscopy techniques offer the potential of imaging molecules in a state closer to physiological conditions, but so far these techniques have only permitted determination of large structures due to the thickness of the layer of ice, = 100 nm. Electron microscopy permits the determination of macromolecular structures which are too large for determination by X-ray crystallography and orders of magnitude too large for NMR. Often when such large structures are studied, useful information may be gleaned at the lower resolutions offered by the electron microscope. Electron microscopy also offers the potential of examining differences in structure of individual members of a population of molecules. In contrast, the other modalities both measure average properties of the sample.
The disadvantages of electron microscopy are its inherent lower resolution compared to the other modalities. This lower resolution becomes more important in light of mutagenesis studies that indicate alteration of single residues in a protein can potentially remove its activity. Electron microscopy would only provide a diagnostic tool to discover links between form and function in such mutations if the change in function was coupled with a large scale change in the structure of the protein, e.g. the inability of the protein to fold to its native conformation.

Finally, there have recently been a number of reports of structures determined by X-ray crystallography being incorrect. Such errors usually occur from incorrect tracing of the protein backbone through the electron density map. Whilst such errors are probably rare, of more concern is the physiological significance of X-ray structures. A recent example that typifies this sort of concern is the structure of insulin. The structure of insulin had been determined by a number of investigators but recently it was discovered that an inactive mutant insulin had a similar conformation. Subsequent NMR studies of both active and inactive forms suggested that the X-ray structure of insulin probably represents an inactive state.

Errors are less likely to be made in the determination of structures from NMR data sets. The only error that may occur during an NMR study would arise if the structure were to adopt a number of distinct conformations. The distance constraints derived from the NMR study would reflect all these conformations simultaneously and thus trying to satisfy such a constraint is equivalent to determining a structure that would probably only be adopted as a transition state. Methods to resolve this problem have been proposed using time-restrained molecular dynamics.

It is seen therefore, that all three modalities have some associated problems. The three techniques of electron microscopy, X-ray crystallography and NMR should be viewed as being applicable in different situations, when different questions are considered. The choice of technique used should always be made in light of the sample under study and of the required resolution.

In the remainder of this introductory chapter, issues pertinent to the electron microscopy techniques used and developed in this thesis will be addressed. In section 1.1.1
the technique of dark-field microscopy will be introduced. Section 1.2 deals with approaches that have been adopted to improve the quality of the input images used in 3D alignment and reconstruction algorithms. This section provides a background to the maximum entropy techniques that are implemented in Chapter 2 and the multivariate statistical approaches employed in Chapters 3 and 4. In section 1.3 the techniques that are available to the electron microscopist for determination of 3D structures are described. Problems and limitations with the various techniques are discussed. This discussion provides a context for consideration of the 3D alignment techniques that are described later in this thesis. The development of the technique described in this thesis was in part aimed at overcoming some of the limitations of previous approaches. However, it also builds on the previous work and, as a consequence, a number of important concepts are introduced in this section.

The work described in this thesis uses the Klenow Fragment of DNA polymerase I as a sample to develop automatic alignment procedures. Previous work on the structure and function of the Klenow Fragment is discussed in section 1.4. Finally in section 1.5 the structure of the remainder of the thesis is described.

1.1.1 Dark-Field Electron Microscopy

This thesis is concerned with the application of electron microscopy techniques to the determination of the structures of globular proteins. While there are many laboratories working on similar problems, the electron microscopy methods and techniques which underlie the novel approaches described below are adopted by very few groups. Perhaps the most fundamental difference relates to the use of a different imaging mode in the microscopy. The majority of workers employ bright-field microscopy techniques in which a combination of amplitude and phase-contrast mechanisms determine the visibility of the object of interest. In the laboratory in which this work was carried out dark-field imaging techniques are employed. Dark-field microscopy is seen to have two related advantages. Firstly, the intrinsic high contrast of dark-field images obviates the need for contrast enhancing agents such as heavy atom stains. The low contrast bright-field imaging mode generally requires the use of such stains and as a result, the resolution of bright-field images is often limited by the size
of the staining agent. Secondly, the contrast in the image is directly related to the mass of the object under study. The relationship between mass and contrast permits the use of computational techniques which rely on interpreting the interrelationships between two-dimensional mass distributions and the three-dimensional structure of the object under study.

Given the apparent advantages of dark-field microscopy, why does bright-field microscopy remain the more popular imaging modality? There are a great many answers to such a question ranging from practical concerns to psychological conditioning. At the practical level, if a resolution on the order of 5 nm is all that is required, e.g., when determining virus structures, then the resolution limit imposed by the staining agent will not be of any consequence. Dark-field imaging of such large (by dark-field standards) structures would also lead to multiple scattering of electrons within the object and thus the relationship between contrast and mass is lost.

For the microscopist used to working in bright-field switching to dark-field requires a fundamentally different approach. For instance, the tight levels on the image screen are on the order of one hundred times less intense in dark-field. In dark-field the image is focused directly whereas in bright-field the in-focus condition of the image results in the loss of all contrast. Similarly, the microscope itself must be aligned in a fundamentally different manner, the apertures required for the two modalities are different and often, as in the case of microscopes without an electronic deflection system, the microscope itself will not be capable of producing high-resolution dark-field images. It should be noted that it is just as hard for a microscopist used to dark-field operation to make the transitions necessary to exploit the advantages of bright-field microscopy.

Thus, the perspective of the dark-field electron microscopist is often quite different from that of the rest of the electron microscopy community. Because the majority of application and technique development centres around the use of bright-field techniques, the dark-field microscopist is often forced into either the development of techniques which exploit the properties of dark-field image formation, or the adaptation of techniques which are more commonly employed in bright-field applications. The work described in this thesis may be considered in this context.
In a wider context, dark-field electron microscopy, an imaging science, is far from being alone. As is the case in many technological fields of science, advances have often arisen from the application of techniques that were developed in relation to other disciplines. Amongst the examples of such cross-disciplinary approaches that are exploited in the work described in this thesis one might include: maximum entropy processing, multivariate statistical methods and the application of quaternions to represent rotations.

1.1.2 The Imaging Environment and Radiation Damage

Observation of high-resolution biological structures using the electron microscope is essentially a destructive process. Potential damage to the structure may occur whilst the specimen is prepared for imaging, and also as a result of the interactions between the sample and the electron beam that occur during the imaging of the sample. These concerns are addressed below in the context of typical procedures adopted for imaging of single molecules without the use of contrast enhancing stains or shadowing, using the dark-field imaging mode.

Generally the microscopist is presented with a solution of the molecule to be studied in an appropriate buffer. Preparation of the sample for microscopy requires the removal of the buffer whilst preserving the structure of the molecule of interest. The simplest dehydration technique is to allow the sample to "air dry" as follows; after application of the sample to the support grid excess buffer and sample are rinsed away with distilled water, or a buffer containing a volatile salt such as ammonium acetate, and the remaining solution is allowed to evaporate under ambient conditions. A concern with the air drying procedure is that surface tension forces may distort the structure of the sample. The alternative procedure, critical-point drying, requires that the sample buffer be replaced with increasingly non-polar solvents, eventually with liquid CO₂ or freon which may be brought through its critical point thus eliminating the detrimental effects of surface tension in the air drying procedure. However, the use of such non-polar solvents may perturb the native structures within the protein under study, (hydrophobic forces are believed to play important roles in defining the tertiary conformations of proteins). To prevent artifacts arising from both of these techniques, cross-linking agents, e.g., glutaraldehyde, may be used in an attempt to maintain the original
conformation of the protein.⁷

As well as concern about the effects of sample dehydration techniques there is also concern about possible interactions between the supporting film and the sample, especially in the context of three-dimensional reconstruction techniques which rely on the molecule retaining a unique structure independent of its orientation with respect to the support film. It has been suggested⁵ that a carbon foil support might be sufficiently plastic as to allow it to wrap around as much as 50% of a specimen. In the same study the authors state from their measurements using small structures (Turnip Yellow Mosaic Virus, ≈ 50 nm) that lack of observed flattening indicates “excellent prospects” for three-dimensional reconstruction. Studies on the 70S ribosomal particles using the carbon film “sandwich” technique⁶ have shown that, at least over a limited angular range, the specimen behaves as a rigid body resting on a support film.⁹

The environment the specimen experiences within the imaging chamber is equally severe. Typically, the sample is in a vacuum of 10⁻⁶ Torr. Stenn and Bahr¹⁰ discuss the effect of such low pressures and conclude that they are likely to have little effect on large molecules but that substances like water may escape from the specimen. Such vacuum-induced loss of water seems to have been largely disregarded as a significant effect by microscopists. This assumption might profitably be reinvestigated in the light of information from other modalities which reveal roles for water molecules as integral components of many protein structures and as having functions in the binding and recognition of substrates.¹¹ The most widely agreed upon effect of imaging in vacuo is that the low pressure will facilitate the escape of fragments of material which result from interactions of the beam and specimen.¹⁰

The alteration of the sample by the imaging beam and the effect of this interaction on attainable resolution has been widely investigated. The sample is subject to an enormous electron flux, estimated¹⁰ to be as high as 10¹⁵ electrons per cm² per second at a magnification of 10,000 x. The quantity of energy absorbed by the sample has not been measured but the transfer of energy from beam to material is described by Bethe’s formula.¹² This relation indicates that the loss of energy is proportional to: the inverse square of the electron velocity (voltage), the atomic number of the constituent elements of the specimen and the thickness and density of the specimen. The quantity of energy absorbed by a biological sample has
been estimated to be of the order of 400 eV per micron per 100 kV electron.\textsuperscript{13} Application of Bethe’s formula may however be inappropriate as it assumes a large irradiated object. Many of the ionization products caused by interaction with the beam are likely to escape from the ultrathin electron microscopy sample, depositing their energy elsewhere.

The primary mechanism for the transfer of energy from the beam to the specimen is a result of inelastic scattering in which electrons in the beam interact with orbital electrons in the sample. Such interactions result in excited or ionized species in the specimen. The excited molecules must then dissipate the excess energy. Pathways available for this energy transfer include emission of electromagnetic radiation (fluorescence or phosphorescence), chemical dissociation, heat production and the formation of chemically reactive ions and radicals. Not all of these processes are necessarily destructive to the sample; most radiation damage is related to secondary events that occur as the induced ions and radicals react with the specimen. At the molecular level the effect of these excited species is the scission of chemical bonds. The breakage is found to be dependent not only on the thermochemical properties of the bond but also on the environment of the bond in the molecule.\textsuperscript{14} In proteins it has been shown that bonds to \( \alpha \)-carbons are particularly sensitive whilst the conjugated bonds of aromatic groups are radiation resistant.\textsuperscript{10} On a still larger scale, the effect of radiation has been shown to be dependent on the class of protein considered. Solubility and aggregation studies indicate that fibrous proteins are more likely to fragment whilst globular proteins are found to cross-link. Tertiary structures of proteins are often stabilized by weak binding forces (e.g., hydrogen, hydrophobic and van der Waals bonding) and that energies of the order 0.01-0.1 eV might disrupt them. In conclusion, it seems that the prediction of the specific forms of radiation damage that might occur within a molecule is not possible unless comparisons with structural analogues may be made.\textsuperscript{15}

What might the effects of such radiation exposures be on the spatial resolution attainable in electron microscope images? The majority of investigations\textsuperscript{9,16-19} which have addressed this question have studied the effect of radiation on electron diffraction patterns of 2D crystalline arrays. These arrays may either be naturally occurring, e.g., purple membrane,\textsuperscript{16} or are artificial 2D-crystals, e.g., L-valine.\textsuperscript{17} Glaeser has developed the following relationship between the dose received by the sample and the expected resolution \( d \),

\[
d = \text{constant} \times \text{dose}^{\text{power}}
\]
where $c$ is the contrast between the object and its surroundings, $k$ is the signal to noise ratio for recognition (generally taken to be 5 after Rose\textsuperscript{11}), $D_{cr}$ is the critical dose required for some endpoint.\textsuperscript{17} Possible definitions\textsuperscript{16} for this endpoint include, the intensity of the diffraction spot corresponding to the highest resolution falling to 1/e of its initial value, or the intensity of the fastest fading spot having fallen by a similar factor.

A wide range of critical doses are found from this type of study. Boudet and Kubin\textsuperscript{20} differentiate between "very radiation sensitive materials" with $D_{cr} < 156 \text{ eÅ}^2$ (electron per Angstrom squared) and "moderately sensitive" material with a critical dose above this limit. (Below a dose of 156 eÅ\textsuperscript{2} they claim that resolution is limited by the modulation transfer function and low sensitivity of the photographic emulsion to 13 Å). From work with purple membrane\textsuperscript{16} it is estimated that the critical dose for a resolution of 7.7 Å is 1 eÅ\textsuperscript{2}. Diffraction patterns of crystals of L-valine are found\textsuperscript{17} to disappear at 0.8 eÅ\textsuperscript{2}, and of adenosine at 6 eÅ\textsuperscript{2}. For uranyl acetate stained catalase stable diffraction patterns have been seen up to doses of 10 eÅ\textsuperscript{2}, at least to a resolution of 17 Å.

Other studies\textsuperscript{21} address the chemical integrity of the specimen after interaction with the electron beam. Electron energy-loss spectroscopy has been used to estimate loss of carbon, nitrogen and oxygen from a variety of materials. The reduction of characteristic ionization peaks was correlated with electron doses, for a 1/e reduction in peak intensity doses range from 14 eÅ\textsuperscript{2} for nitrogen from collodion to 249 eÅ\textsuperscript{2} for oxygen from graphite bisulphate. Loss of carbon was not seen.

The doses required for dark-field microscopy at magnifications typically used for high-resolution studies of proteins are typically much greater (800-1500 eÅ\textsuperscript{2}) than the critical doses determined in the above studies. However, it has been demonstrated that protein substructures\textsuperscript{22-24} on the order of 5 Å and heavy atoms\textsuperscript{25,26} may be visualized directly using dark-field techniques. Whilst the assignment of resolution to these studies which are based on subjective, repeated recognition criteria is difficult, independent confirmation of the
validity of the observations has come from the subsequent determination of the same structures using X-ray crystallography.  

To explain the visibility of such high-resolution structures in the context of the critical doses given above, further consideration should be given to the results of radiation damage. The retention of carbon at high electron doses has been proposed to leave a stable predominantly carbon structure which may be what is observed. Other concerns exist with regards to the extrapolation of the results from the diffraction studies, particularly that the studies cannot distinguish between the loss of long-range order in the crystal and damage occurring within the unit cell.

It has also been postulated from modelling studies that even extensive damage to a structure may not necessarily be detectable except at resolutions of around 0.3 nm. Similarly modelling radiation damage as a random walk phenomenon indicates that alpha-helical structures might be visualized at a resolution of 3.2 Å after a dose of 844 e Å².

The use of cryo-electron microscopy techniques that keep the specimen at liquid helium temperatures will also reduce the effect of radiation damage in the image. Damage to the specimen is still thought to occur but the low temperatures prevent diffusion of molecular fragments by what is referred to as the “caging” effect. Cryo-electron microscopy has increased the critical dose, Dc, by up to a factor of 5.

1.2 Image Restoration and Enhancement

High-resolution electron microscope images often contain significant amounts of noise. Restoration and enhancement techniques are commonly applied to such images to permit the extraction of the useful structural information from the noisy image. The algorithms that are used for such processing fall into two broad categories: those in which the algorithm acts solely to remove the effect of the degradative noise processes; and those in which the aim is to extract some parameter or feature information from the image. The latter group of algorithms, which includes techniques such as contouring, edge enhancement, histogram equalization and thresholding, are generally non-linear techniques and as a result the processed images are inappropriate as input images for 3D reconstruction procedures.
Image enhancement nomenclature and theory has followed the pre-existing conventions of signal processing. Generally the recorded image is considered to be composed of the "true" uncorrupted image convolved with an instrumental response functions plus an additive noise component. The aim of the algorithms described below is the recovery of the uncorrupted image from the recorded noisy image.

There are many sources of degradation that affect the quality of electron microscope images. Theoretically the noise in a dark-field electron image should arise primarily from the electron counting statistics of image formation. In practice other random processes dominate the degradation of the image. Perhaps the single largest degradation of the image results from the carbon film which supports the object of interest. Variations in the thickness and scattering of this support film are reflected as an additive noise signal in the image.

The most straightforward image enhancement technique relies on the principles of the central limit theorem and consists of the summation of images that are of putatively identical objects which are degraded by noise.\(^{33-35}\) Summation of \(N\) images should result in a signal-to-noise ratio enhancement by a factor of \(\sqrt{N}\). In electron microscopy, the use of averaging techniques is complicated by the translational and rotational freedom of the object of interest. If the images that are summed do not represent identical views of the same object then the resolution of the resulting average image will necessarily be lower than that of the individual images. In section 12.2.2, techniques for the identification of similar images will be discussed more fully.

Averaging principles also underlie the signal-to-noise ratio enhancements that result from the use of 2D crystalline arrays of molecules.\(^{19,36}\) Once the lattice spacings of the crystal have been determined the average content of a large number (typically thousands) of unit cells may be determined using Fourier techniques, only the structures of the individual images that repeat with the same periodicity as the lattice will make a contribution to the ensemble average. When the object of interest possesses either rotational\(^{37,38}\) or translational symmetry\(^{39}\) this may often be "enforced" in composite images and the images' quality enhanced. Another related approach\(^{40}\) employed optical filtering techniques to elucidate the average structure using a pseudo-crystalline array of images that were aligned visually with respect to one another.
Enhancement of images of single molecules has also been achieved by the use of filtering techniques. Although there are many possible image transformations\textsuperscript{41} which permit such filtering, the most common is the Fourier transform.\textsuperscript{42} Fourier transformation allows images to be considered as a summation of independent spatial frequencies. The underlying assumption in the use of Fourier filters for signal to noise ratio enhancement is that the characteristics of the signal and the noise, in terms of their spatial frequencies, will be different from one another. Because the Fourier transform of an image allows processing of individual frequencies, those which are thought to constitute noise in the image may be removed. Selective spatial filtering techniques, or band pass filtering, may be used to remove high frequency noise which shows up as speckle in the image and low frequency variations in the image (e.g., due to differences in illumination levels across the image or to long-range variations in the characteristics of the support film). Care must be exercised, however, in the use of high-frequency filters. If the high-frequency cutoff is set too low, the resulting image will appear blurred, because the high spatial frequency features in the object of interest will also be removed. The assumption that the signal and noise components of the image are separable in terms of their frequency components is also usually not true and as a result spatial frequencies may be lost from the object of interest if spatial filters are not used correctly.

Another approach that relies on the use of Fourier transformation of the image is “optimal” or Wiener filtering. In this approach the image is assumed to have been corrupted both by convolution with a blurring function and by the addition of a noise component. The form of the optimal filter may be determined if it is possible to differentiate between the relative contributions of the underlying blurred function and the noise to the frequency distribution in the resulting image. The optimal filter may then be applied to the image followed by deconvolution of the blurring, to return the uncorrupted image.

The potential loss of high-frequency information within the object of interest when band-pass filtering is applied led to the investigation of other image enhancement techniques. In the following section maximum entropy methods for image restoration will be discussed. This section provides a context for the consideration of the application of maximum entropy methods to dark-field micrographs which is described in Chapter 2. In section 1.2.2 the use
of multivariate statistical techniques to identify similar images which may then be summed to determine ensemble averages will be examined. Multivariate statistical techniques are employed to facilitate image enhancement of both simulated images and dark-field electron micrographs later in this thesis.

1.2.1 Maximum Entropy Methods

At the commencement of the work described in this thesis the maximum entropy method had been reported as being a useful data analysis technique in a large number of fields including X-ray astronomy,\textsuperscript{43} \(\gamma\)-ray astronomy,\textsuperscript{44} optical deconvolution,\textsuperscript{45} the solution of structures from X-ray crystallography studies\textsuperscript{46,47} and NMR spectroscopy.\textsuperscript{48} General purpose algorithms for image enhancement had been described by a number of authors.\textsuperscript{49-51} The theoretical justifications of the method and its demonstrated success in the above applications suggested that the application of maximum entropy techniques might be of use when applied to the problems of restoration of electron microscope images.

A mathematical description of the principles of the maximum entropy method are given in appendix 1 of Chapter 2. The principles of the method may be stated as follows. Suppose the measured data consists of an observed quantized frequency distribution. On the basis of this measurement we wish to estimate the most likely form of the distribution that the measured data represents. If the distribution is governed by purely random events, then the most likely outcome will be the one which could be realized in the largest number of ways, i.e., the outcome with the highest multiplicity. Use of Stirlings' approximation shows that the multiplicity of an outcome may be directly related to the entropy of the distribution, 

\[ -\Sigma p_i \ln p_i, \]

where \(p_i\) represent frequencies in the distribution. Thus, the distribution which maximizes the entropy is identical to the frequency distribution which can be realized in the greatest number of ways. An alternative justification of the method notes that the above definition of entropy is equivalent to minus the Shannon information content of the frequency distribution.\textsuperscript{52} The maximum entropy distribution is thus the distribution that contains a minimum \(\Delta\) information whilst remaining consistent with the measured frequency distribution. Consistency between the calculated maximum entropy distribution and the
measured frequencies is usually enforced by the use of a $\chi^2$ constraint.

The principles of the maximum entropy method as currently applied rely on the work of Jaynes$^5$ and arose from his work on estimating prior probabilities. The work of Shore and Johnson$^4$ and Tikochinsky et al.$^5$ are commonly cited$^6$ as evidence of the rationality of the use of the maximum entropy as a selection criterion. Tikochinsky et al. illustrate through the use of simple examples that consistency requirements (e.g., repeating an experiment should not change its result) will only be satisfied through the use of a maximum entropy selection criterion.

Initially the problems that have been tackled by the application of maximum entropy methods in astronomy appear to be closely related to the image enhancement problem in electron microscopy. Closer investigation reveals some differences, most notably that the image is usually collected in Fourier space in astronomical applications and often the image will be degraded by convolution with the point spread function of the observing instrument. The point spread function in the microscope does not play a role in limiting the resolution. The noted differences suggest that the electron microscopy image formation process is in fact more closely related to the probabilistic arguments given for the maximum entropy criteria for image selection, than the astronomy applications.

In many applications of the maximum entropy method different entropy expressions to that given above are employed.$^{50,57}$ One example, referred to as relative entropy,$^5$ is written $\sum p_i \ln(p_i/m_i)$ where $m_i$ represents some estimate or model of the distribution. The argument for the inclusion of a prior model is as follows. If there are no constraints on the reconstructed data then the maximum entropy reconstruction will be a uniform image with the mean data intensity. Constrained maximum entropy reconstructions from noisy data will be biased towards the mean of the data. If the true image is close to the mean then this may not present a problem. However this situation is the exception. The use of relative entropy will in turn bias the maximum entropy solution towards the model distribution described by the $m_i$.

In Chapter 2 experiments which examine the use of maximum entropy methods applied to the image enhancement problem in dark-field electron microscopy are described. The implications and consequences of the use of modified entropy expressions is discussed.
further in Chapter 2. In the application of maximum entropy methods to electron microscopy images described in Chapter 2, the more conventional entropy expression is used.

The results described in Chapter 2 indicate that application of simple maximum entropy algorithms leads to solutions that are not consistent with the form of the noise in the original data. In Chapter 2 this problem is overcome by adding novel constraints to ensure that the noise is removed from the image in a physically reasonable manner.

1.2.2 Pattern Recognition Techniques Applied to Image Averaging in Electron Microscopy

1.2.2.1 Introduction

The principle of enhancing noisy images of single molecules through the averaging of similar images remains one of the most predominant in electron microscopy.9,38,62 As indicated above, image averaging techniques are only useful when, (1) the images do indeed contain identical views of the molecule, (2) the structures to be enhanced are in register (rotationally and translationally) with one another, and (3) the images differ only as a result of random noise in the images, i.e. there are no systematic variations in the image population. Because the original electron microscope images are usually extremely noisy, large numbers of images must be combined to generate three-dimensional structures of useful resolution. Visual alignment and classification of such large image sets is neither practicable nor objective. The rapid advances in computer technology in the 1970s and '80s which generated machines capable of storing and processing extremely large amounts of data enabled the development of many powerful automatic alignment and classification algorithms. The following sections will review these various algorithms. Approaches to the alignment of images prior to classification are discussed in section 1.2.2.2. The data compression technique of principal component analysis as applied to electron microscope images is discussed in section 1.2.2.3. A number of different image classification strategies are discussed in section 1.2.2.4.

1.2.2.2 Automated Image Alignment

Procedures for the automatic alignment of images of single particles were, in general, developed for application to bright-field images of large macromolecular complexes.63,64 Such
complexes were often found to adopt a limited number of orientations with respect to the support film and thus present a limited number of projections in electron micrographs. Because "characteristic" views of the molecule could often be recognized visually, one such view was usually suitable as a reference against which the other images in the data set were aligned using rotational and translational cross-correlation functions. The alignment was then refined using the average of the ensemble of aligned images of the previous cycle as the new reference. Cross-correlation techniques have become the standard methods for determination of relative alignments although other techniques such as the stochastic sign change method described by Bonnet and Liehn have been suggested.

Often a molecule will adopt a number of distinct orientations with respect to the support film. In such cases a better alignment of similar images will be achieved if they are compared to a reference image derived from their own population rather than with respect to a global average. To address this issue a multi-reference alignment scheme has been implemented. The multi-reference alignment procedure aligns the images with respect to a single reference image as before, the images are then classified using the multivariate statistical techniques described below. Averages are formed from each of the classes and are used as references to inter-compare the population of images. The images are then aligned with respect to the reference to which they are most similar (on the basis of their maximum cross-correlation coefficients).

The logical problem with both of the alignment schemes described above is that they require the choice of an initial "good" reference image and that the choice of such an image relies on the knowledge of the structure under study, which will usually not be known. The dependence of the methods on a high quality reference image was illustrated by Boekema et al. who showed that the relative alignment of images was sensitive to the choice of reference image. Furthermore when the images are extremely noisy, features in the noise that resemble the reference will reinforce and as a result the average of the aligned images will resemble the initial reference image.

Recently a reference-free alignment procedure has been described by Penczek and Frank. Following a discussion with Penczek, but prior to the publication of their algorithm, the author developed a similar procedure. Despite being developed separately the algorithms
are essentially the same (differences are noted in section 4.1.1 of Chapter 4). The algorithm is an iterative procedure in which images are sequentially aligned with respect to one another as follows. The second image, in a data set of many images, is aligned with respect to the first. The aligned images are then summed to form a composite reference image. The third image is aligned with respect to the composite reference and, once aligned, it too is added to the composite image. This process is repeated for each of the images in the data set. Following a "first pass" the contribution of each image is removed from the composite and the image realigned with respect to the remainder of the composite. The algorithm proceeds until no positional changes occur during realignment. Penczek shows that the algorithm will in fact converge to an optimum alignment due to the non-negativity of the cross-correlation functions used to determine individual, optimum alignments. However, it is likely that the optimum found will only be a local optimum because the image rotations are only considered individually.

The reference-free alignment technique has been used by Penczek and Frank to align the quasi-continuous distribution of images of the 70S ribosome embedded in ice. In this thesis the algorithm is applied to the alignment of heterogeneous populations of both simulated and real dark-field electron microscope images of the Klenow Fragment.

1.2.2.3 Multivariate Statistical Analysis

In the following section a mathematical framework that permits the recognition and classification of similar images will be discussed.

An image may be regarded as a point in an $N$ dimensional hyperspace, where $N$ is the number of pixels in an image. In the experiments described below $N = 4096$. The coordinates of the point in the hyperspace are given by the intensity values of the pixels. If we consider $M$ images, then they will form a cloud of $M$ points in the hyperspace. As stated above $M$ will often also be large, of the order of 1000. Two images will be considered similar if they are close together in this hyperspace, i.e., the intensities in each of the pixels are similar. In determining which images are similar, and thus should be averaged, we are essentially looking for groupings, or sub-clouds, of images in the hyperspace. To determine exactly the optimal grouping, or classification, of images in such a large dimensional space is a
computationally intractable problem, the number of possible grouping is at least of the order of $M!$. The techniques of multivariate statistical analysis, specifically principal component analysis, have been employed to reduce the dimensionality of the image classification problem.

Principal component analysis characterizes the distribution of images in a new, lower dimensional coordinate space by describing the cloud of images in the $N$ dimensional space in terms of its principal components (also referred to as factors or eigenimages). The first axis of the new coordinate system represents the direction of greatest inter-image variance of the population of images. The second axis, which by definition will be orthogonal to the first, represents the direction of the largest remaining variance of the image population, and so on. Often relatively few (10-15) eigenimages need to be calculated to characterize the important features of the population of images. The eigenvalues (corresponding to each eigenimage) describe the contribution of each eigenimage to the variance of the population. Having established the new coordinate system, the positions of the $M$ individual images may be calculated in the new lower dimensional hyperspace by projecting their positions on the new coordinate system. It is assumed that the largest contributions to the inter-image variance of the population of images will arise from differences in orientation of the molecules and thus the coordinates in the lower dimensional space will reflect these differences in orientation, allowing the images to be classified in the lower dimensional space. However, there will also be contributions to the inter-image variance from effects such as conformational flexibility of the molecule, structural deformations and, in stained preparations, variability in the distribution of the stain. Image noise, from the statistics of the imaging process and variability in the support film, will normally only be reflected in the higher order eigenimages which are not considered in the classification procedure. The neglect of the higher order eigenimages in subsequent classification procedures may in itself be considered as an image enhancement technique.

The decision about the number of eigenimages that should be employed in subsequent classification steps has until recently been made either on the basis of inspection of the relationship between the eigenimage number and the variance it accounts for, or recognition that certain eigenimages represent features which are to be separated in the subsequent
partition. Recently Harauz and Chiu\textsuperscript{79} have applied the technique of event covering to determine the statistical significance of the different eigenimages and to select the most relevant eigenimages that will be used in classification algorithms. The technique evaluates the interdependence of, (1) the coordinates of each image in the hyperspace described by the eigenimages and (2) class membership. In applying this technique to simulated images\textsuperscript{79} and real images\textsuperscript{80} they describe the recognition of non-significant eigenimages and claim that event covering provides a more relevant description of the relative importance of the eigenimages in describing class structures than the eigenvalues themselves.

1.2.2.4 Classification

The use of principal component analysis techniques described in the previous section will reduce the dimensionality of the classification by a factor on the order of 200. The aim of classification procedures is to determine the intrinsic structures of the cloud of points in this lower dimensional space and to recognize subgroups, or classes, of images that are similar enough so that they will, on summation, yield a high resolution image with an increased signal to noise ratio. Such a class of images may be defined from a number of perspectives.\textsuperscript{77} The "geometric" perspective suggests that an image should be considered to be a member of a class if it is closer to the centre of that class than to any other. The "probabilistic" interpretation of a class is, a region in the hyperspace that is densely populated with images but is surrounded by a zone which is sparsely populated. Just as there are a number of definitions of what constitutes a class of images, so there are many methods for trying to determine them, the merits of the different methods arousing considerable controversy in the literature.\textsuperscript{81,82} What is apparent, however, is that there are no clustering or classification criteria that are universally applicable or superior, that selection of a criterion will be subjective and that no classification method will be optimum for all data sets.\textsuperscript{82}

There are two principal approaches, referred to as hierarchical and partitional methods, that have been adopted to solve the classification problem.\textsuperscript{83} The application of the hierarchical method (which is commonly employed in biological taxonomy) to image processing was first suggested by van Heel.\textsuperscript{73} This method relied on the theorem which states that: the total variance of the set of data points is equal to the sum of the inter-class variances.
plus the total intra-class variances for all the classes. An optimum partition is thus suggested as one in which the inter-class variance is maximal. In order to facilitate classification the proposed algorithm used the criterion of "minimum added intra-class variance" to decide which two images or classes of images should be merged to form a new class. This simple classification procedure can lead to non-optimum partitioning of the data, however, and a modification of the procedure was implemented to allow a degree of re-partitioning of the data after the initial hierarchical classification.

A hybrid partitional/hierarchical method has also been proposed. This algorithm combined a $K$-means-type algorithm with a hierarchical classification scheme. In $K$-means algorithms the euclidean distance between data points and the centres of gravity of each class are calculated, and on the basis of this distance the data point is assigned to the centre of gravity of the nearest class. The algorithm is "seeded" with random class centres and the classification process ceases when no further changes in the partition are seen.

The decision about the optimum number of classes in all the studies described above was based on examination of the hierarchical trees (or dendrograms) which illustrate the merging of classes and the resulting increase in intra-class variance. The assumption being that a large increase in intra-class variance indicated that two different populations of images were being merged. While this method is intuitively sensible it has been ranked only 11th when compared to 30 other techniques applied to the problem of determining the numbers of classes in data sets analyzed by a variety of hierarchical methods.

Recently Carazo et al. have applied the concept of "fuzzy sets" to the classification of electron microscopy data sets. The use of fuzzy sets was suggested as they possess the ability to represent grades of membership within a class. The results indicate that fuzzy sets clusterings were in agreement with other methods previously used and that a number of fuzzy set cluster-validity functionals proved to be useful in determining if the data possessed a clustered structure.

A slightly different approach was adopted by Borland and van Heel. They compared the classification of images described in "pixel space" to the classification of pixels in "image space". Classification in the somewhat less intuitive conjugate space enabled the identification of pixels of the image which exhibit similar behaviour within the data set of images.
comparing the classification analyses in the two lower dimensional spaces it is possible to
determine which features in the image are important in defining a class.

It should be noted, that while the above discussion considers the techniques most
widely employed, a different approach to the classification problem has been proposed by
Schatz and van Heel. Their method, based on rotationally and translationally invariant
autocorrelation functions of the images, obviates the need for the initial alignment step prior
to classification. This approach (and the problems associated with it) is discussed more fully
in Chapter 4. It is also noted that no reference to the practical application of this method is
found in the literature.

1.3 Three-Dimensional Structures from Electron Microscopy

The following sections examine the contributions that electron microscopy has made to the
elucidation of 3D biological structures. Many different approaches have been adopted to
facilitate the determination of such structures, the choice of approach often being determined
by the properties of the specimen. In reviewing the application of electron microscopy to
structural determination problems it is convenient to consider each of the approaches
separately. The principles underlying each of the techniques will be discussed along with
eamples of the results, associated problems and limitations. Although considered separately,
the methods will be seen to share many common features, use similar principles and share
computational techniques. The methods are reviewed in no particular order but the techniques
considered will generally become more similar to the technique which is described later in
this thesis.

1.3.1 Analysis of 3D Structure through the Visual Interpretation of Electron Micrographs

The most straightforward approach to the determination of 3D structure using the electron
microscope is simply to deduce the structure from observation of features in the 2D electron
micrographs. While this approach might seem somewhat subjective it has often proved to be
quite successful. Determination of 3D structures from a single 2D projection image is
theoretically impossible. However, examination of many images or the incorporation of other knowledge about the object under study will often suggest a structure which is consistent with both the observed images and the supplementary information. This approach to structural determination is much less rigorous than those described below but it has the advantage of being simple to perform. The obvious disadvantage is the subjectivity involved in the interpretation of the images and their relationship to the other available data.

One example of a 3D structure deduced from the interpretation of electron micrographs is that of protamine, a 4 kD molecular weight DNA binding protein. On the basis of analysis of images with a resolution of 0.5 nm in the context of knowledge of the sequence of the protein, coupled with physical and chemical data, a 3D structure consisting of a curved loosely helical structure was postulated. This interpretation of the images as a representation of the 3D structure was later confirmed when the X-ray crystallography structure became available.

A similar analysis was possible for glucagon, a 29 amino acid polypeptide. On the basis of high-resolution electron micrographs and knowledge of the sequence of the molecule, it was possible to postulate a 3D structure. The proposed structure was different from highly helical structure determined by X-ray crystallography, but was in agreement with the other biological and chemical data, probably representing an alternative conformation of the molecule.

While the secondary structure of small proteins may be proposed from analysis of electron microscope images, model structures from larger proteins are necessarily lower resolution. Working on larger particles, Andrews et al. were able to postulate a structure for the signal recognition particle (M.W. 226 kD) on the basis of bright- and dark-field micrographs and the use of heavy metal shadowing which provides visual cues as to the proportions of the molecule. Analysis of the images led to identification of a subunit structure within the molecule. Subsequent use of electron energy loss imaging techniques allowed determination of the location of the nucleic acid within the structure. Energy loss imaging has also been used to permit a low-resolution interpretation of images of 7S ribonucleoprotein particle composed of one molecule 5S RNA and one of the 40 kD molecular weight DNA binding protein TFIIIa.
The four molecules described above provide examples of the sort of 3D information that may be determined from high-resolution electron micrographs. This sort of analysis results primarily from the expectation that proteins will be three-dimensional in nature, highly two-dimensional structures being uncommon in biology. The sort of analysis described above probably forms the preliminary step in all the subsequently described techniques; in order to decide what approach to adopt, the microscopist must generally make some preliminary observations about the structure that is being studied.

A slightly different interpretational approach is possible when the structures of related molecules have been determined previously. In such cases, knowledge of the related structure may be used to interpret the new images. This sort of approach has been used to understand the relationship between different subunits of the ribosome\textsuperscript{97} and to determine that the structure of the ribosome is apparently conserved across some species.\textsuperscript{38}

An interpretational approach has also been possible when a molecule occupies a number of distinct orientations with respect to the support film. Comparisons of the features of different views will often suggest a single consistent 3D structure. Such an analysis was possible for haemocyanin\textsuperscript{74} and a chaperonin oligomer.\textsuperscript{37} Both of these molecules presented two types of projections, “top” views containing some symmetrical elements and “side” views that were readily related to the “top” view by a rotation of 90° perpendicular to one of the axes of symmetry.

1.3.2 Determination of the Three-Dimensional Structure of Symmetrical Objects

The presence of symmetry in a biological object presents the electron microscopist with the possibility of obtaining a large amount of information about the 3D structure of the object from a single, or limited number of views. This possibility was first exploited by De Rosier and Klug\textsuperscript{38} in 1968 when they were able to reconstruct the tail of the bacteriophage T4 to a resolution of 3.5 nm. The tail consists of a helical repeating structure. Thus, intensity cross-sections through the tail at different positions along its length are equivalent to a series of cross-sections through a single section of the tail at different rotational orientations. De Rosier and Klug recognized this fact and were able to employ Fourier synthesis techniques
to incorporate the cross-sections in a 3D structure.

An extension of this method came when the group of Crowther applied similar principles to the reconstruction of the Tomato Bushy Stunt virus and the Human Wart Virus. Instead of using a series of 1D sections as in the previous example, 2D images were combined to reconstruct the 3D object. The principle was that objects possessing a 3D symmetry would present similar views at a variety of orientations. Therefore it was necessary only to collect one, or a few views, to perform a 3D reconstruction.

Both of the methods described above are based on the Projection of Central Section Theorem. The theorem may be stated as follows: the 2D Fourier transform of a projection of a 3D density distribution is identical to the corresponding central section of the 3D Fourier transform of the same distribution normal to the direction of the view. A similar relationship exists between 2D density distributions and 1D projections. The central section theorem is fundamental to the work described later in this thesis and it will be discussed at greater length in Chapter 3.

The original 3D reconstructions of De Rosier and Crowther were performed using images of specimens stained with heavy atoms and visualized with bright-field microscopy techniques. The procedure that is more commonly used today for the determination of virion structures is that of cryo-electron microscopy in which the specimen is embedded in a layer of vitrified ice approximately 100 nm thick. The major advantage of this technique is that there is no requirement for resolution-limiting stain or for fixatives to maintain the structure of the large virions in the microscope. Also the entire virion density will be imaged such that 3D reconstruction techniques will reveal not only the external structure but also any ordered internal structure. Using cryo-electron microscopy techniques the 3D structures of a large number of icosahedral viruses have been determined. Despite the comparatively large size of these structures, (the largest being the Herpes Simplex Virus MW 450 MD, with a diameter of 125 nm), it is often possible to achieve resolutions on the order of 2 nm.

An interesting variation on this approach appeared recently in the literature. The group of Baker were able to visualize antigen binding fragments (Fabs) bound to the surface of Cow Pea Mosaic Virus to a resolution of 2.3 nm. The success of their approach suggests similar approaches to examining other structures. It may be possible to bind structures to viruses, via
an antibody intermediary molecule, and then to exploit the symmetry of the virus to provide a high quality reconstruction.

Cryo-electron microscopy techniques have also been applied to molecules possessing helical symmetry. The sheath of the 10 nm diameter fibre of F-actin has been described recently. The resolution achieved was such that individual residues could be located via the use of gold labelling techniques.\textsuperscript{39}

The structures that have been derived from symmetrical objects benefit from averaging that may be performed over the individual cells which make up the symmetrical unit as well as from precise knowledge of relative projection directions. The relationship between these techniques and those described later in this thesis was indicated by De Rosier in 1968 when he postulated that the central section theorem might be applied to reconstruct asymmetrical objects if the relative orientations of different projections could be determined. The technique described in this thesis will use the central section theorem itself to determine these orientations.

1.3.3 Three-Dimensional Reconstruction Using Tilt Series

Determination of the 3D structure of an object from projections requires knowledge of the orientations of the projections with respect to one another. Perhaps the simplest resolution of this problem is to collect a series of projection images at known orientations. The use of a tilting specimen support stage and goniometer makes this possible. A series of images is collected with the specimen at known orientations with respect to the electron beam; this series is referred to as a tilt series.

There are a number of problems associated with the use of tilt series. One of the most severe is related to the requirement for multiple exposures which results in a large dose to the sample (see below). A second problem is that a full range of views may not be collected. Normally the maximum attainable tilt is around ±45°, although some stages may be capable of ±80°.\textsuperscript{111} Even using high-tilt stages problems are encountered due to the overlap of specimens at high tilt and the increased effective object thickness. The limited angular range of views leads to what is referred to as the “missing cone problem”, a region of the
reconstructed volume in frequency space for which there is no data. The net effect of the missing cone is that the reconstructed object will have poorer resolution in the direction normal to the tilting axis than in the direction parallel to it. Mathematical approaches have been proposed to try to estimate the missing cone data with some success.\textsuperscript{112}

Despite these problems many biological objects have been reconstructed using this method. The pioneering work in this field was carried out by the group of Hoppe.\textsuperscript{113-115} Their first reconstruction was of the 2.3 MD protein Fatty Acid Synthetase, which was reconstructed using 9 projection images. The molecule which has a diameter of 35 nm was reconstructed at a resolution of 1 nm parallel to the imaging plane and 3-4.5 nm in the direction normal to this. Similar studies have been performed on the 50S and 30S ribosomal subunits.\textsuperscript{116,117} In all cases the images were taken using bright-field techniques with negative staining. The total dose in some of the ribosome reconstructions, which used 21 exposures over a ±60° range was 1250 eÅ\textsuperscript{-2}.

The use of 2D crystalline arrays allows the high dose levels to be reduced because the dose is distributed over the crystalline lattice. An averaging of elements of the lattice may be performed using the Fourier filtration techniques discussed in section 1.2. Structures of molecules from both natural and artificial crystalline arrays have been derived using tilt series analysis.

The work by the group of Henderson is perhaps the most well-known application of tilt series techniques to natural 2D crystals.\textsuperscript{118} They recently published the structure of Bacteriorhodopsin at a resolution of 0.35 nm parallel to the imaging plane. The resolution was sufficiently high as to enable visualization of the α-helices which make up the core of the protein, of the bulky aromatic residues attached to these helices and of the β-ionone ring of the retinal chromophore. Combining tilt series, image processing techniques to correct lattice distortions, averaging techniques and cryo-electron microscopy this structure perhaps represents the gold-standard of electron microscopy derived structures to date.

Although there are other structures that form naturally occurring 2D arrays,\textsuperscript{119} the majority of proteins do not. However, it is sometimes possible to grow 2D crystals of proteins and then use tilt series to generate 3D structures. Sometimes crystals may be grown simply from the molecule of interest,\textsuperscript{120,121} in other cases, exemplified by the study of cytochrome...
oxidase by Valpueta et al., vesicle membranes may be used to facilitate the formation of a 2D crystalline array. The applicability of combining tilt series and 2D crystals is limited, however, by the difficulties experienced in trying to obtain high quality 2D crystals.

1.3.4 The Random-Conical Tilt Method

As mentioned in the previous sections one of the principal problems with the use of tilt series for 3D reconstruction is the high doses that are required to obtain images of the molecule in many different orientations. One method to avoid this problem is the use of 2D crystalline arrays described above. A second method has been proposed by Radermacher and is referred to as the random-conical tilt method. The method requires two images of a field of identical molecules, only one of which is used for the 3D reconstruction. The method relies on the molecule adopting a preferred orientation with respect to the plane of the support film with random in-plane orientations. The first image of the sample is taken at a tilt of 50°, the second is untilted. The untilted image is used to determine the azimuthal angle (in-plane rotation) of each molecule. This information and the images of the molecule in the titled image may then be combined in a reconstruction algorithm as if the views had been acquired from a conical tilt series.

Because only a single micrograph is required, the dose to the sample is around 0.3 - 0.4 eÅ⁻². The method has been successfully applied to images of the ribosome and its subunits. The resolution between independent reconstructions of the 50S ribosomal subunit derived from the same data set is around 4 nm. A recent extension of the method by Carazo and Frank permits the incorporation of views of a molecule that still exhibit preferred orientations but are able to “rock” about an axis parallel to the support plane. This technique is referred to as multicone reconstruction.

The random-conical tilt method has proved to be a useful method for the determination of the structure of asymmetrical objects using low electron doses. However, its application is limited by the requirement that the specimen must adopt a few recognizable preferred orientations with respect to the support film.
1.3.5 Reconstruction of Randomly Oriented Asymmetrical Particles

Determination of the 3D structure of a molecule that lies in random orientations with respect to the support film has proved to be extremely difficult. However, there have been a number of methods proposed to solve this problem.

The group of Frank was able to solve the problem by interpreting the images from a single micrographs as a representation of a continuum of views of the molecule. Multivariate statistical analysis was performed and the images were partitioned into different classes on the basis of their coordinates with respect to the 5th and 6th eigenimages. It was then possible to calibrate the interconversion between classes of images as a rotation by comparing the images to those obtained from separate tilt series of similar images, and by comparison with pre-existing models. The resolution of the resulting reconstruction was estimated to be 2-3 nm. This reconstruction method is dependent on the molecule lying in orientations corresponding to rotation about a single axis or upon recognition of such a population of images from a set of images in a truly random distribution of orientations.

Another approach to the problem was that described by Harauz and Ottensmeyer. They were able to reconstruct the nucleosome core particle by using a pre-existing model of the core particle to assign preliminary orientations for 3D reconstruction. The proposed super helical coil of DNA was compared to the phosphorus signal obtained from electron energy loss images of the core particle to estimate the projection directions. Once the initial angles had been assigned they were refined using a self-consistency algorithm which compared pseudo projection images of the reconstructed volume to the original electron micrographs. Using this technique the estimated resolution of the reconstruction was 1.5 nm.

Whilst these methods were successful in determining 3D structures, in both cases it was necessary to have additional information about the structure under investigation. The following section considers possible approaches to the 3D reconstruction problem that are based entirely upon information contained within the projections.
1.3.6 Previous Approaches to Automatic Orientation Determination

Determination of the 3D structure from randomly oriented projections is an attractive technique for two reasons. Firstly because there is no requirement for either special preparation techniques or a tilt stage. The second advantage is that it should be possible to determine the 3D structure from a single micrograph of many particles; thus the dose requirement for this technique is similar to that required for the random-conical tilt procedure and is significantly less than for tilt series. To exploit these possibilities the orientations of the molecules must be determined from the projections themselves. A number of authors have proposed and demonstrated methods which permit such determinations.

The use of the central section theorem (see section 1.3.2) for a posteriori determination of projection directions was proposed by van Heel\textsuperscript{128} in 1987 and by Goncharov and co-workers around the same time.\textsuperscript{129} Both groups describe the method in real space and Goncharov also formulates the equivalent Fourier space solution. The proposed solution of the alignment problem may be stated in real space as follows. The central section theorem states that any two 2D projections of a 3D density distribution will share a common 1D line projection. Once the positions of the common lines have been determined they define axes about which the pairs of projections are constrained to rotate. If three projections are intercompared then the axes of rotation fix their relative orientations, and a system of equations may be defined which determine the orientations of the projections with respect to one another. The method is referred to by various names including the common axis method or common line method; van Heel refers to the techniques as angular reconstitution.

In the paper by van Heel, he describes the use of sinograms, basically collections of line projections at 1° intervals, and their sinogram correlation functions, as a method for determining the positions of common line projections. He demonstrates the technique via application to a simple model for three projections. He gives no indication of the sensitivity of the method with respect to noise.

The implementation described by Goncharov is essentially similar to that of van Heel. The method was demonstrated using images of a density distribution resembling that of the ribosome and projection directions corresponding to three faces of a tetrahedron. Noise was
added to the projections, the value of the noise corresponding to 15 to 30% of the maximum density value in the image. Line projections were calculated at 10° intervals. The projection directions were determined correctly by the algorithm. However, this result is somewhat related to the choice of projection directions and line projection sampling intervals.

While these two papers demonstrated the feasibility of the common axis method neither of them adequately addressed the potential of the method when applied to real data. In addition, the described geometrical methods for the determination of projection directions discuss only the relationship between three projections. The method in Fourier space was discussed by Goncharov, and later by Van Dyck, but in neither case was a practical implementation described.

The common axis method was proposed to determine the projection directions of a large number of randomly oriented particles. However, the method will work just as well with specimens that occupy a few characteristic orientations with respect to the specimen support. This situation is perhaps preferable because averaging techniques may be used to improve the quality of the input images for the technique. The method will fail, however, if the particle possesses certain symmetries or if the orientations of the molecule are related by rotation around a single axis, as is often the case for elongated particles.

The use of moments of the mass distribution of an object to determine its orientation has been proposed by a number of authors. In 1988 Goncharov and Gelfand proposed the method as a technique for avoiding problems that occur when common axis methods are applied to particles which are rotated about a single axis parallel to the support plane. In their paper they showed how second and third central moments of an image (considered as a 2D mass distribution) may be related to moments of the 3D mass distribution of the particle. They show that $N$ projection images will yield $7N$ equations in $9+3N$ variables ($9$ moments of the original distribution and $3$ Euler angles for each projection). They noted that this system is “rather complicated” and did not present any method for solution. Instead they reduced the scale of the problem by considering the case of an object rotating about a single axis. Thus there is only one angle to determine for each projection. They presented a numerical search algorithm which is able to locate a solution. Using model data (four different sized spheres) the algorithm was shown to work for a limited number (<10) of
projections and a signal to noise ratio of 2, provided that the other Euler angles did not vary by more than 10° in the projections. The same authors presented another solution to the same problem of determining coaxial rotations by examining 1D projections of the moments in sections normal to the axis of rotation. They claimed this second technique is somewhat more stable than the former technique.

An attempt to solve the more general orientational problem via the use of moments was presented by Salzman in 1990.\textsuperscript{132} The method relies on sampling a large number of projections of objects in random orientations. This allows statistical determination of the principal moments of inertia of the 3D mass distribution. Orientations of projections are then determined by comparing the moments of the 2D mass distribution in the image to the principal moments of the 3D distribution. Because the principal moments (eigenvalues of the inertia matrix) are known, the problem of determining the rotation matrix becomes an inverse eigenvalue problem, i.e., given the diagonal elements of a matrix, reconstruct the full matrix. Salzman presents a method for solving this problem. Using a model density distribution, Salzman plotted the variation of two second order moments against Euler angle and indicated that the moments will specify the orientations. The paper is however rather short on specifications as to the success of the method. He estimates that a 1% increase in the noise level in the image should increase the "error in establishing alignment" by 2%.

The advantage of the method over common axis techniques is a reduction in computing time of the order of the square root of the number of pixels in the image. The disadvantage of the method is the need for sampling of each of the principal moments, implying the need for a large number of input images. Not sampling the principal moments will introduce bias into the assigned projection directions.

In conclusion, while a number of techniques for the determination of orientations of randomly oriented particles have been proposed, none of them have been particularly well characterized and thus the relative merits of the methods are difficult to determine from a simple review of the literature. The procedure that is adopted in this thesis is based on the common axis method. The thesis describes the extension of the method to the alignment of many projections and characterizes the behaviour of the technique when applied to realistic model data and real electron micrographs.
1.4 The Klenow Fragment of DNA Polymerase I

The experimental studies described in this thesis centre on investigations of the 3D structure of the Klenow Fragment of DNA Polymerase I. The 3D alignment technique that is described later in this thesis could be applied to reconstruct the 3D molecular density distribution of many different molecules. The reasoning behind the choice of the Klenow Fragment is explained in the following sections.

1.4.1 Previous Studies

DNA and RNA polymerases occupy a central role in biology. The survival of a biological system is dependent on the ability of that system to accurately replicate the nucleic acid genome. Because of their similarity of function it might be expected that DNA and RNA polymerases are both evolutionarily and structurally related and that as a result, knowledge of the structure of DNA polymerase may suggest the form of prebiotic polymerases. Studies of DNA polymerases have concentrated on two issues, what role does the enzyme play in assuring the fidelity of template directed DNA synthesis and, how does the polymerase enzyme process along the DNA. The answers to both these questions will obviously be intricately linked to the structure of the polymerase and so a great deal of effort has been directed towards such structural studies.

The first DNA polymerase was isolated by Kornberg in 1958 from *Escherichia coli*. It consisted of a single polypeptide with a molecular weight of 103 kD. It was originally called simply DNA polymerase, but following the subsequent discovery of other polymerases it was renamed DNA polymerase I. DNA polymerase I is involved primarily in the repair of damaged DNA and the processing of Okazaki fragments. DNA polymerase II (molecular weight 120 kD) is believed to play a role in repair of DNA damage. DNA polymerase III (molecular weight 140 kD) is responsible for the replication of the chromosomal DNA. In all cases the polymerase is just one of several proteins that are required for replication.

DNA polymerase I has three enzymatic activities, a DNA polymerase, a 3'-5' exonuclease activity which serves to edit out mismatched nucleotides at the 3' end of the
extended strand and, a 5'-3' exonuclease activity that removes the DNA ahead of the growing strand of DNA. The enzyme has separate binding sites for deoxynucleoside monophosphate (dNMP) and triphosphate (dNTP). Enzymatic activity and nucleotide binding require divalent ions, probably magnesium in vivo.

The three enzymatic activities of DNA Polymerase I have been found to reside in three separate domains of the molecule. Proteolysis yields two fragments, the 35 kD N-terminal domain that retains the 5'-3' exonuclease activity and a larger 68 kD fragment that contains both the polymerase and 3'-5' exonuclease activity. The large fragment, or Klenow Fragment as it became known, was cloned into an expression vector by Joyce and Grindley. This made large quantities of the molecule available for biochemical and structural studies.

In 1985, the X-ray crystal structure of the Klenow Fragment was solved to a resolution of 0.33 nm by the group of Steitz. The structure revealed the molecule to have two domains. The small N-terminal domain of approximately 200 amino acids consisted of a core of β-pleated sheet flanked by two α-helices. This domain binds two divalent ions and a molecule of dNMP. The larger carboxy-terminal domain forms a structure that contains a large cleft, about 0.20-0.24 nm wide and 0.25-0.35 nm deep. The bottom of the cleft is formed by a six-stranded anti-parallel β-sheet; α-helices form its sides. The structure has been compared to that of a right hand grasping a rod. Thus, one wall of the cleft, the fingers, consists of 6 disordered α-helices and is 5 nm long. The other side of the cleft, corresponding to the thumb is considerably smaller, formed by just two α-helices. At the top of the thumb there is a section of 23 amino acids that were not resolved in the crystal structure and are presumably disordered or have a flexible attachment to the rest of the molecule. See Fig. 1.

The polymerase activity of the molecule has been located to the larger domain of the Klenow Fragment, primarily by synthesis of the carboxy-terminal alone, followed by the observation that it retained the polymerase activity (at a reduced level). The carboxy-terminal retained no exonuclease activity. Furthermore, cross-linking studies indicate the dNTP binding site was located at the end of one of the α-helices which protrude into the cleft. The recent co-crystallization studies of the Klenow Fragment with double and single stranded DNA by Freemont et al. reveal, as was postulated, that the cleft in the large domain does
Figure 1. The structure and main features of the Klenow Fragment of DNA Polymerase I as determined by X-ray crystallography. The tubes represent α-helices, the ribbons represent β-sheets.

form a binding site for the double stranded DNA.

That the smaller subunit contained the 3'-5' exonuclease site was suggested by the observation that it bound dNMP, which is known to inhibit exonuclease activity. Site directed mutagenesis studies within the smaller fragment have also been able to remove the exonuclease activity. Finally the co-crystallization study showed that the single stranded DNA bound within the smaller subunit.

The intriguing observation is that the exonuclease and polymerase active sites are separated by a distance of 3 nm. Joyce has demonstrated that the Klenow Fragment may exhibit both activities without requiring dissociation from the DNA. However, this mode of action is not always used and the DNA may move from one active site to the other via
dissociation from the DNA. The choice of pathway is governed by the relative rates of dissociation from the DNA and the polymerase and exonuclease activities. It has been shown that these rates are dependent on whether the terminal DNA is correctly or incorrectly paired and upon the local DNA sequence.

The mechanism for the interconversion between polymerase and exonuclease modes was suggested by Freemont et al. on the basis of their co-crystallization studies. They determined that the DNA would have to slide 8 base pairs between the two sites and that a melting of 4 base pairs would be required to bring the single stranded 3' end to the exonuclease site. It is also postulated that access to both active sites may require the 3' terminal to slide between the two sites. Part of the evidence supporting this is the observation that the 3'-5' exonuclease activity removes approximately 10% of correctly paired nucleotides. This translational mechanism of action is also supported by the chemical kinetics work of Dahlberg and Benkovic.142

1.4.2 Selection of the Klenow Fragment for Electron Microscope Studies

From the previous work described above, it was apparent that the relationship between the structure and function of the Klenow Fragment presented an interesting problem. For this reason alone the study of the Klenow Fragment by the use of dark-field electron microscopy techniques was suggested as being a fruitful area of research, not only to confirm the X-ray structure previously determined but to possibly identify conformational differences between the two modalities and correlate them with conformational freedom of the molecule.

There were a number of factors that suggested that the Klenow Fragment would make a particularly good test molecule for the application of the techniques described in this thesis. The size of the molecule, its molecular weight of 68 kD and dimensions (10, 8 and 6 nm) suggested that identification of the molecule in dark-field images would present no serious problems. The molecule possessed no symmetries which would present a problem to common-axis-based alignment techniques. The X-ray structure does, however, exhibit features at a range of scales, the overall dimensions of the molecule, the large and small subunit structure, the DNA binding pocket and smaller features such as the unresolved amino acids

36
Figure 2. Examples of the similarities between dark-field electron microscope images of the Klenow Fragment of DNA polymerase I (rows 1 and 3) and density projections of the structure derived from X-ray crystallographic studies (rows 2 and 4). Bar represents 5 nm. Note that even though the images are all displayed at the same scale, the molecule in the X-ray derived projections appears slightly smaller. The difference in size is interpreted in part to be a result of only the α-carbon positions being used to generate the simulated images.

at the top of the thumb that were possibly resolvable in the electron microscope images. The preliminary observation of electron images indicated that features of the image could be interpreted in terms of the X-ray crystallography structure. See Fig. 2.

Another important factor was the availability of the X-ray structure, deposited in the Brookhaven database. Unfortunately only the positions of the α-carbons were available from the 0.33 nm resolution structure, but it was possible to extrapolate this data and combine it
with known properties of the amino acids to make pseudo-electron density projection images of the molecule. These simulated images could then be used to test the automatic orientation assignment algorithms. Use of a similar structure to the experimental images was important as the success of the alignment algorithms will be dependent on the amount of structure present in the molecule (see discussions in Chapters 3 and 4).

After the work described in Chapter 4 was completed, we were able to obtain from the Steitz group an electron density surface of the Klenow Fragment. This allowed comparison of the features of the electron microscopy models to that of the X-ray derived structure with greater confidence than when using the model derived from the α-carbon data alone.

The Klenow Fragment was also chosen for study because it offers future possibilities for study bound to its DNA substrate in a similar fashion to the co-crystallization studies of Steitz. Application of electron energy loss imaging techniques also offer the exciting possibilities of locating the DNA in the complex. Recently a study of the Klenow Fragment with and without DNA bound was performed using scanning tunnelling microscopy. The resulting images show some evidence of a two domain structure, although it bears little resemblance to the X-ray crystallography structure. The authors interpret the images as indicating a change in conformation on binding of DNA.

1.5 Structure of the Thesis

The overall aim of the work described in this thesis was the development of techniques that permit the 3D reconstruction of molecular density distributions from dark-field electron microscope images. At the outset of this work it was apparent that improvement in the quality of the microscope images would lead to improvements in the 3D reconstructions. To this end maximum entropy methods were investigated as a possible image enhancement procedure. This work is described in Chapter 2. Application of maximum entropy techniques in a conventional manner was found to introduce artifacts into the images. Chapter 2 describes how, by using additional constraints, this problem was overcome. However, because the algorithm that results is rather impractical the method was not considered further. Readers primarily concerned with the development and application of 3D alignment techniques may
wish to initially skip to Chapter 3. The work described in Chapter 2 may then be considered at a later time in light of some of the more successful image enhancement techniques employed later in this thesis.

Chapter 3 describes an algorithm that permits the assignment of relative orientations of a large number of projections images *a posteriori*. This algorithm is central to all the subsequently described simulations and experiments. In Chapter 3 the principles of the algorithm are discussed in detail. Simulated data is used to indicate the usefulness of the technique. Following on from this work, Chapter 4 describes the application of the algorithm to simulated dark-field electron micrographs of the Klenow Fragment. The work described in Chapter 4 was undertaken to establish that the algorithm could cope with the sort of images likely to be obtained from the electron microscope. Chapter 4 describes a number of image processing techniques which are subsequently applied to images from the electron microscope. The work with simulated images led to the development of a slightly modified algorithm, capable of more accurate determination of projection directions when presented with noisy images.

The results of Chapter 4 were sufficiently encouraging for us to pursue structural studies of the Klenow Fragment using dark-field electron micrographs. These studies are described in detail in Chapter 5. The practical aspects of working with digitized electron micrographs are considered and adaptation of previously discussed image processing techniques are described. The results of the electron microscopy based study were somewhat disappointing, especially in light of the previously described simulations. A possible explanation for the poor performance of the algorithm is given at the end of Chapter 5.

Chapter 6 provides a more general discussion of the relationship between the results of the simulations and those obtained using electron micrographs. A number of possible methods for improving the performance of the algorithm are discussed. Also in Chapter 6, the results of two brief studies are described. One of the studies indicates the relationship between one of the algorithm parameters (used throughout Chapters 3, 4 and 5 to indicate the performance of the algorithm) and the quality of the 3D reconstruction one might expect. The second study considers the 3D reconstruction that results when images that are visually selected as being of high quality are used in alignment determination/3D reconstruction.
process. The results of the latter study are somewhat more encouraging than those described in Chapter 5.

Chapter 7 summarizes the contribution of the work described in this Thesis to 3D structural studies using dark-field electron microscopy.
References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Maximum entropy methods are examined for the removal of noise from dark field electron microscope images. A number of maximum entropy algorithms, including $\chi^2$ and $E^2$ constraints, are examined. Failure of conventional maximum entropy methods led to the inclusion of an autocovariance constraint in the algorithm. Under this constraint the noise removed from the images has a more realistic spatial distribution than under conventional constraints. This constraint, when combined with a simulated annealing algorithm, led to successful removal of noise from small test data sets.

Published in shorter form in *Ultramicroscopy*, 31, 275-284 (1989)
2.1 Introduction

Interpretation of minimal dose electron micrographs taken at high resolution in dark-field or bright-field microscopy is often complicated by the presence of large amounts of noise in the images, which obscures the features of interest. The noise arises from a variety of sources. By far the largest contribution results from the imaging of the structure of the specimen support, usually a thin carbon film. Strictly this noise is stationary in time but its random spatial details can preclude the recognition of other structures placed upon it. The second source of noise is time varying or exposure varying details which have as their source the radiation induced structural alterations of the specimen and the support film. The third source of noise is quantum noise, the effects of which become significant as exposures are reduced in order to minimize structural alterations due to the electron beam. Electron imaging thus has to be a compromise between attempts to minimize radiation damage on the one hand and the use of sufficient numbers of electrons to overcome the effects of quantum noise and permit the recognition of the object of interest at a desired resolution on the other. The problem that we address in this paper is the design of an image processing algorithm to allow the extraction of useful structural information from the noise within a single image.

Averaging techniques, taking advantage of the redundancy of information in crystalline arrays,\(^1\) or techniques such as low pass filtering\(^2\) have been used to address the noise removal problem. For a single image however noise reduction has only been achieved at the expense of a reduction in the spatial resolution of the image.

Maximum entropy methods have been employed in a variety of fields to remove noise from data sets. The methods are apparently most successful in the fields of nuclear magnetic resonance spectroscopy\(^3\) and in radioastronomy image deconvolution.\(^4\) Both of these problems are classical “inverse problems” in which noisy data is normally convolved with an instrument measurement function which results in the data becoming blurred. Gull and Skilling\(^5\) have reviewed the application of maximum entropy methods to such inverse problems. In applying maximum entropy methods to the problems of noise removal from dark-field electron microscope images we are considering their application to a fundamentally
different problem. Examination of $\chi^2$ and $E^2$ constrained maximum entropy algorithms applied to this problem indicated either only slight enhancement of the noisy images or the introduction of artifacts. However, the $E^2$ algorithm became more pertinent to our problem after the introduction of an autocovariance constraint on the noise removed from the image.

2.1.1. The $\chi^2$ Constrained Maximum Entropy Method

The illustration and theory commonly used to justify the maximum entropy method is applicable not only to inverse problems but also to the noise removal problem in dark-field electron microscopy. It is conventional to justify the acceptance of maximum entropy methods by considering the probabilities of different distributions arising when balls fall randomly with equal probabilities into an array of bins. It can be demonstrated that the most likely distribution is the one which maximizes

$$S(F) = -\sum_{i=1}^{k} f_i \ln f_i$$

where $f_i$ is the probability of a ball landing in box $i$. $S(F)$ is known as the entropy of the distribution $F = \{f_i: i=1, k\}$. Following such an argument it is asserted that, in the selection of a single image from the set of images that fit the data but have reduced noise, the “best” image is the image which maximizes expressions of the form given in eq. 1.

Conventionally a $\chi^2$ measure between a trial image $F$, and the measured data is used to constrain the selection of the image with respect to the data. For the above probabilistic argument to be valid it is required that the sum of the image pixels be unity; thus all data and image values must be scaled to allow them to be considered as probabilities. In appendix 2.A.1 the probabilistic arguments for the use of maximum entropy arguments are described. A description of the algorithm used to determine the distribution $F$ that maximizes the expression in eq. 1 under a $\chi^2$ constraint is also given.

The results of applying a $\chi^2$ constrained maximum entropy algorithm to a simulated dark-field image are shown in Fig. 1. The maximum entropy image, Fig. 1b, does appear less noisy than the original and the absolute value of the noise standard deviation has decreased; but so has the contrast in the image. However, when the image is displayed at a different
Figure 1. A simulated dark-field image and its maximum entropy image: (a) the simulated image; (b) the maximum entropy image displayed at the same contrast levels as (a); (c) the maximum entropy image displayed with maximum contrast. Comparison of (a) and (c) reveals no significant enhancement arising from maximum entropy modelling.

contrast level we see that both the structure of the data and the noise in the image are virtually unchanged. Application of the algorithm to a 1D data set containing a rectangular signal is more directly interpretable. Figure 2 illustrates that the main effect of applying the maximum entropy method is to produce a non-linear compression of the data (Fig. 2b).

The noise used in the generation of the data in Fig. 2 had a Gaussian distribution. The distribution of the residuals, the differences between the data and the maximum entropy image shown in Figs. 2a and 2b, has been plotted in Fig. 2c. The distribution is bimodal, obviously non-Gaussian. It is strongly affected by structures in the data. This is clearly unacceptable. If the maximum entropy methods are to remove the noise from the data one would expect that the distribution of removed noise (the residuals) be the same as that present in the data – Gaussian in this example.

2.1.2 *The E^2 Constrained Maximum Entropy Method*

Application of the method of error fitting originally described by Bryan and Skilling allowed for correction of some of the problems associated with the *χ^2* constraint. This method
Figure 2. The 1D data and maximum entropy image resulting from the application of a $\chi^2$ constrained maximum entropy algorithm to a noisy rectangular function: (a) the data; (b) the maximum entropy image; (c) the distribution of residuals between the data and the maximum entropy image. Residual size is expressed in standard deviation of the noise in (a).

replaces the $\chi^2$ constraint on image selection with the error fitting ($E^2$) constraint. The latter compares the residuals to an expected noise distribution rather than comparing the data and the maximum entropy image directly. A formal description of the $E^2$ constraint and its incorporation in a maximum entropy algorithm is given in appendix A.2.

The $E^2$ algorithm was applied to a similar data set as that used to test the $\chi^2$ algorithm. The results are shown in Fig. 3. The distribution of sizes of the residuals by virtue of the $E^2$ criterion is now forced to be similar to the noise distribution. Nevertheless, whilst the distribution of the residuals has been fitted correctly, two artifacts have been introduced in the maximum entropy image. These are a shift in local means in the maximum entropy image and a skewing of the noise in different regions of the image.

These artifacts arise for two reasons. The maximum entropy method seeks an image as close as possible to the global mean of the data, and the spatial distribution of the residuals is not constrained to be the same as that of the noise in the original data.

A further difficulty is associated with the method of error fitting. The ordering step required to allow comparison of the residuals to the expected noise distribution makes the $E^2$ constrained entropy function discontinuous and only differentiable between ordering steps.
Figure 3. The 1D data and maximum entropy image resulting from the application of an $E^2$ constrained maximum entropy algorithm to a noisy rectangular function: (a) the data; (b) the maximum entropy image; (c) the spatial distribution of the residuals between the data and the maximum entropy image. Residual size is expressed in standard deviation of the noise in (a).

Each subsequent reordering of the subtracted noise array alters the end point sought by the algorithm. In spite of this problem, convergence to a high entropy solution is achieved due to the gradual change in the order of the residuals between iterations.

In this paper a new algorithm is described that finds maximum entropy images constrained not only by the size distribution of the noise removed but also by their spatial distribution by use of an autocovariance constraint. In addition, methods of simulated annealing are used to avoid the problems associated with the discontinuous nature of an $E^2$ constrained entropy function.

2.2 Simulated Annealing Applied to the Entropy Maximization Problem

Simulated annealing has been successful in providing solutions to a number of problems known as combinatorial minimization problems. These problems require a minimum to be found in a function which is defined not in a space of $k$ continuously varying parameters but in a discrete, usually very large, configurational space. The method of simulated annealing has been used to solve very effectively the most famous of the combinatorial problems, that
of the travelling salesman. In this application its use avoids the problems of ordering and the consequent discontinuity of the entropy function associated with the error fitting algorithm.

To apply simulated annealing to the selection of a maximum entropy image we must couple some of the ideas used in the error fitting algorithm with those of simulated annealing. The image is now generated by the subtraction from the original data set of an array of expected noise values. In subtracting such an array, the condition that the size distribution of the noise in the data and the noise removed from the data be the same will automatically be satisfied. We equate the spatial configuration of the noise array that is being removed with the configuration of the crystal lattice in the natural manifestation of the annealing process. The entropy of the image that results from the subtraction of a particular configuration of residuals with the data is equated to the energy of the crystal lattice. For \( k \) pixels in the image there are \( k! \) possible ways to combine the noisy data with expected noise arrays. Simulated annealing allows us to search efficiently to find the optimum combination.

The physical annealing process allows stable structures to be found because the probability that an element of the system is at a given energy \( E \), is given by the Boltzmann probability distribution

\[
\text{Prob}(E) \propto \exp\left(-\frac{E}{KT}\right)
\]

As a result, a slowly cooling system does not necessarily get trapped in local minima because there is the possibility that as the system cools the energy of a part of the system may increase.

To find the optimal configuration, the algorithm proceeds in the following manner. Rather than searching through all the configurations, the merit of sequential random reconfigurations is examined. Two types of random rearrangement are made to the noise to be subtracted from the data. The first type of rearrangement involves the reversal of a random length section of the array at a random position in the array; the second involves the translocation of a random length section of the array to another randomly chosen position in the array. Having made these rearrangements, the entropy of the image resulting from adding the data and noise arrays is recalculated. The entropy of the image is compared to the entropy of the previous image. The difference between the two entropies is referred to as the entropy
Figure 4. The 1D step function data presented to the simulated annealing algorithm and the maximum entropy solution image produced. The unbroken line represents the step function underlying the noisy data and also the solution that the algorithm is designed to recover.

The cost of rearrangement, $\Delta S$.

We must introduce the concept of a temperature for the system which will allow us to set up the equivalent of a Boltzmann distribution. The distribution used is defined

$$Prob(\Delta S) \propto \exp(\Delta S/T)$$

The temperature $T$ is an algorithm parameter that necessarily has units of entropy. As it is reduced to simulate cooling, it will control the acceptance of rearrangements. If the entropy is higher after rearrangement, $\Delta S$ is positive, and the new configuration is accepted. If the cost is negative, then the chance of acceptance of the new configuration is governed by the above distribution. The value of $T$ is initially selected such that every rearrangement is accepted. The algorithm makes $10^k$ successful rearrangements, the temperature is then lowered by 10%. As the temperature falls the chances of making successful rearrangements also falls. The algorithm will, at each temperature, search for $10^k$ successful rearrangements before the value of $T$ is reduced. The value of $T$ is also reduced if, after $100^k$
rearrangements, 10 $k$ successful rearrangements are not found. The process is complete when no successful rearrangements are found in 100 $k$ trials.

To find the one exact solution among the $k$ factorial possible configurations, quickly becomes impractical as $k$ increases. Therefore, in order to allow us to study the applicability of simulated annealing in comparison to the search through all configurations for the exact solutions, we examined a simple step function of only 10 points, to which Gaussian noise was added. The maximum entropy solution found by simulated annealing is illustrated in Fig. 4. For only 10 points we are able to calculate the entropy for every possible combination of noise residuals with the data, and thus determine the exact configuration yielding the entropy maximum. The solution found by the method of simulated annealing agrees with this exact determination. The calculations for 10! combinations required 2.2 CPU hours on a MicroVax II. The method of simulated annealing made 54485 trial rearrangements, of which 6716 were accepted; this required 212 CPU seconds on the same computer. It can be seen in Fig. 4 that in applying the method of simulated annealing as described above, the artifacts that were found in the solutions using the error fitting constraint remain. The application of simulated annealing in the presence of a spatial constraint is discussed in section 4.

2.3 Introduction of a Spatial Constraint

In the Introduction we described artifacts in $E^2$ constrained maximum entropy images. These images are constrained only by their sum and by the size distribution of the residuals. In Fig. 3c we have plotted size of the assigned residuals against position for the data set and maximum entropy image shown in Figs. 3a and 3b. It is clear from Fig. 3c that the spatial distribution of the residuals is influenced by the structure of the rectangular signal in the original data. The noise in the original data is not spatially distributed in this manner. Locally, in either the background or in the signal, the noise mean is zero.

We have used the autocovariance function to examine the spatial properties of these distributions. The normalized autocovariance functions of the rectangular function, of the noise added to it to produce the data, and of the residuals, are plotted in Fig. 5. We apply similar arguments to the spatial distribution of the residuals as were used to justify the $E^2$
constraint; if the noise is to be removed from the data to yield the rectangular function then
the spatial distribution of the removed noise must be the same as the spatial distribution of
the noise that was added to produce the data. It is apparent from examination of Fig. 5 that
this is not the case when the error fitting algorithm is used. The form of the autocovariance
function of the removed noise, Fig. 5c, and the autocovariance function of the noise added
to form the data, Fig. 5b, are quite different.

This constraint may be incorporated into the maximum entropy image selection
process. It is required not only that the distribution of the noise is known, but also that its
spatial distribution must be characterized in terms of its autocovariance function, either
theoretically or experimentally. The autocovariance between array values separated by \( r \), or
lag \( r \), for the residuals \( U = \{ u_i; \ i=1, k \} \) is defined

\[
ACVF(r) = \frac{1}{k-r} \sum_{i=1}^{k} u_i u_{i+r}
\]

An autocovariance function is similarly defined for the noise that was added to the
rectangular function to form the data. A constraint term is then set up by comparing the
autocovariance function of the noise present in the data, \( ACVF^* \), and the residuals between

---

Figure 5. The autocovariance functions of: (a) the noise-free rectangular function in
fig. 3a; (b) the noise present in the data in fig. 3a; (c) the residuals between the data
and the maximum entropy image shown in fig. 3c. Note the dissymilarity between (b)
and (c).
the data and the maximum entropy image, $ACVF^*$, using the following measure to quantify the differences between the two autocovariance functions

$$\kappa = \sum_{r=1}^{k-1} (ACVF^*(r) - ACVF^-(r))^2$$

### 2.4 Inclusion of an Autocovariance Function Constraint

Incorporation of the spatial constraint in the simulated annealing entropy maximization image selection process is achieved by redefining the cost function. Instead of defining the cost of rearrangements in terms of entropy alone, a new cost $\Delta S'$ is defined as

$$\Delta(S') = \delta(S) - \mu \Delta \kappa$$

where $\Delta \kappa$ is the difference in the parameter $\kappa$ defined in eq. 5 resulting from the rearrangement of the residuals. The parameter $\mu$ controls to what degree the solution is influenced by the autocovariance constraint. If $\mu$ is set to zero then an image is selected only on the basis of maximizing the entropy under the error fitting constraint. If $\mu$ becomes large, however, the $\Delta \kappa$ term becomes more important and selection of the image will be influenced by the similarities between the spatial distribution of the residuals and the noise. The annealing process will no longer select the global maximum entropy image but will select the maximum entropy image that has a spatial distribution of residuals that fits that of the noise to a degree determined by $\mu$.

Figure 6 shows the distribution of $\kappa$ and entropy for all the possible images that could be generated from the data shown in Fig. 4. On this figure the locus of the values of $\kappa$ and of entropy has been plotted for images generated by the method of simulated annealing using different values of $\mu$. The figure reveals that as expected the algorithm has always selected the highest entropy image available for a given fit between the autocovariance functions. The image that is returned by the algorithm for values of $\mu$ equal to $10^3$ is the noise-free step function (the unbroken line in Fig. 4). The value of $\kappa$ for this image is zero. In principle there are always two solutions which have a $\kappa$ of zero, one in which the noise is removed perfectly and one in which the noise is added in a manner identical to the original noise in the data.
Figure 6. The distribution of \( k \), the fit between the autocovariance functions and the entropy of the image for all possible images generated from the 1D data shown in fig. 4. The locus of the simulated annealing solutions for different values of \( \mu \) is indicated by the arrows.

The algorithm is able to select between these two images on the basis of their entropy.

2.5 Further Considerations

2.5.1 The Application of Simulated Annealing

The method of simulated annealing is one of the most computationally efficient ways of solving combinatorial maximization/minimization problems. The speed of the algorithm described above is proportional to \( k^2 \). As a result we soon encounter prohibitive calculation times, a 256-point problem of the type described above required 31 CPU hours on a MicroVax II. The computational time can be reduced by a factor of \( k/\log_2 k \) by using fast Fourier transforms to calculate the autocovariance functions. This reduction alone will not
yield computing times fast enough for practical implementation of the above simulated annealing algorithm. A further improvement can be gained by careful consideration of the algorithm annealing schedule, i.e. the initial temperature and the rate of cooling.

The application of simulated annealing to optimization problems is a relatively new technique, one for which no general formalism yet exists. The method has been applied to a diverse set of problems, the travelling salesman, hard disc phase transitions, single photon emission computer tomography and potential energy surface searching. In most of these examples it has been necessary to adapt the general principles of the simulated annealing method so that they pertain to the particular problem under investigation.

We chose an implementation of simulated annealing which most closely paralleled that of Kirkpatrick. The values of the parameters involved in the annealing schedule i.e. the temperature, the rate of cooling and the numbers of rearrangements examined were chosen from practical experience. Because we initially chose to work with a small data set (10 points) there was no great imperative for us to optimize the algorithm with respect to speed of convergence.

Geman and Geman have considered the problem of image restoration by using simulated annealing to search for a maximum a posteriori estimate of the image when the original model is based on a Gibbs distribution. They provide a theorem leading to an annealing schedule which guarantees convergence to the global maximum of the posterior distribution. Such a schedule states that the temperature $T(j)$ at the $j^{th}$ iteration should satisfy

$$ T(j) \geq \frac{c}{\log(1+j)} $$

for every $j$, where $c$ is a constant independent of $j$. However the values of $c$ that were determined by the authors lead to a prohibitive computing burden that could only be overcome by substituting arbitrarily higher values of $c$. The logarithmic cooling described by eq. 7 allows the algorithm to search for greater times at low temperatures (close to freezing).

Szu and Hartley claim to have developed a "fast simulated annealing" algorithm that cools the system at a rate inversely proportional to the number of iterations. This is made possible by adapting the reorganization of elements of the array and by making both the reorganization of the array and the acceptance criterion temperature dependent. Thus at higher
temperatures, more large scale rearrangements are attempted than at lower temperatures near the freezing point. Their approach is demonstrated in both one dimension\textsuperscript{14} and \( n \) dimensions.\textsuperscript{15}

2.5.2 Other Entropy Expressions

It should be noted that the entropy expression we have used to determine the entropy of an image is not the only entropy expression that is commonly used in the literature. Often prior information will be incorporated\textsuperscript{16} yielding an entropy expression of form

\[
S_A(F,M) = -\sum_{i=1}^{k} f_i \ln(f_i/m_i)
\]

where the \( m_i \) are the estimates for the distribution \( F \). Another form of entropy used\textsuperscript{17} is defined

\[
S_B(F,M) = \sum_{i=1}^{k} \left[ f_i - m_i \right] m_i \ln(f_i/m_i)
\]

where \( S_B \) is the entropy of the image relative to the model \( m \). It is the case for both these expressions that they are maximal when image \( F \) is equal to the model \( M \).

We have two problems with using these entropy expressions. Our first problem lies in the choice of \( M \). The least biassed choice of \( M \) is one in which each pixel value is the mean of the data \( P \). In using our entropy expression (eq. 1) we effectively use this model. If one uses a low-pass filtered version of the data then any maximum entropy image will automatically tend towards such an image, not a desirable feature as the resolution in the final image will inevitably be less than that of the data. We have examined the behaviour of an algorithm using entropy of the form given by eq. 8 with both low-pass and median filtered images as models for the maximum entropy images. This work leads us to our second concern: both these models, lie within the \( \chi^2 \) or \( E^2 \) surfaces. As a result no tension is set up between the entropy and the constraint. The filtered model would thus be chosen as the maximum entropy image.
2.6 Conclusions

For any given noisy data set there are many images that are consistent with data and have less noise in them than the original data. The problem facing us is to select one of these images as being the best. Following the traditional arguments leads one to select the maximum entropy image, the “best” in the sense described in the Introduction. However we have shown that this “best” image is in general seriously flawed and that the maximum entropy constraint leads to biased non-linear results that have to be viewed with great caution.

Maximum entropy images selected using a $\chi^2$ constrained algorithm provide no significant enhancement of signals in test models of noisy dark-field images. The main reason for this lack of enhancement is that the image which is “best” as understood by having a high entropy is the image that is closest to the global mean of the total data set. This is an inappropriate “best” image for any situation where the local mean varies with position in the image, i.e. a dark-field image or a bright-field image with amplitude contrast. The applicability for pure phase contrast bright-field images, remains to be determined.

When maximum entropy images for the type of data set modelling dark-field images are examined it is clear that the noise removal process is unrealistic in both its intensity distribution and its spatial distribution. Constraining both the intensity distribution with an $E^2$ constraint and the spatial distribution with a constraint on the autocovariance of the residuals allows the maximum entropy method to produce a more realistic image. These constraints are sufficient for the return of the correct image when the algorithm is applied to a very simple step function for which the noise and the spatial distribution of the noise were known exactly.

In real processing situations the accuracy of the final maximum entropy image will be governed by the accuracy of the numerical descriptions of the known parameters in the image. In this chapter the image noise was used as an example, and as the most objective known parameter. Any further knowledge of the image or the objects imaged can be incorporated as further constraints. In this growing list of constraints the maximum entropy term becomes a minor though in the end still helpful component.

The method of simulated annealing was shown to be capable of finding solutions to
constrained maximum entropy problems when applied to small data sets. For implementation of the method with much larger data sets, further optimization of the annealing schedule will be required. It is likely that by exploring further the parallels between the thermodynamic processes and simulated annealing, greater insight will be gained as to the choice of parameters leading to the optimization of the annealing schedule. Extrapolating from present results, an improvement of algorithm speed by a factor of at least $k/\log_e k$ must be obtained for the practical processing of images with $k$ pixels.

Acknowledgements

The authors thank Todd Constable for his useful discussions in the course of this work. Neil Farrow is a Research Student of the National Cancer Institute of Canada. This work was supported by grants from the National Institute of Canada, the Medical Research Council of Canada and the Ontario Cancer Research and Treatment Foundation.
Appendices

2.A.1 The $\chi^2$ Constrained Maximum Entropy Algorithm

The image is generated by $N$ electrons falling into a photographic plate which is then digitized into $k$ zones or pixels. The image produced will contain two components, the component of interest, which results from structures in the sample, and the noise component, which will manifest itself as a random distribution of electrons distributed throughout the whole image. Consider the image as a frequency distribution. If we have $n_i$ electrons falling into an area covered by pixel $i$ then we can assign to that pixel a frequency $p_i$ given by

$$p_i = \frac{n_i}{N}$$  \hspace{1cm} (A.1.1)

The image as a whole is then considered to be a collection of frequencies $P = \{p_i; i = 1, k\}$. Assuming that a purely random process governs the distribution of electrons in the pixels, the likelihood of a given set of frequencies is expressed by the multinomial coefficient or multiplicity factor $W(P)$ defined

$$W(P) = \frac{N!}{n_1!n_2!...n_k!}$$  \hspace{1cm} (A.1.2)

Using Sterling's approximation

$$\ln(a!) = a(\ln(a) - 1)$$  \hspace{1cm} (A.1.3)

it can be shown that

$$\frac{1}{N} \ln W(P) = -\sum_{i=1}^{k} \left( \frac{n_i}{N} \right) \ln \left( \frac{n_i}{N} \right)$$  \hspace{1cm} (A.1.4)

$$= -\sum_{i=1}^{k} p_i \ln p_i$$  \hspace{1cm} (A.1.5)

The term on the right-hand side of the above equation is, by definition, the entropy of the distribution of the set of probabilities, denoted $S(P)$. The image most likely to be
generated by the random process is the one that has the largest multiplicity factor. From eq. A.1.5 this image is also the distribution with the maximum entropy.

For any given data set containing noise there will be many model images, or sets of frequencies \( F = \{ f_i; i = 1, k \} \), which are consistent with the data. Maximum entropy theory demands that from all these models we must select the one which maximizes the entropy of \( M \) as defined by eq. A.1.5. Consistency between the data and model can be established by the use of a statistical test. The metric that is most commonly used to constrain the model with respect to the data is \( \chi^2 \) defined as

\[
\chi^2 = \sum_{i=1}^{k} \frac{(p_i - f_i)^2}{\sigma^2}
\]

in which \( \sigma^2 \) is the standard deviation of the noise in the original data set. The expected value of \( \chi^2 \) is \( k \).

A second constraint that must be applied to the selection of the model is that the sum of the frequencies in the model must be equal to 1. This is required if we are to use the preceding multiplicity maximization justifications for the method.

To find the model that has the maximum entropy, the method of Lagrange multipliers is used. Formally, the function to be maximized is defined

\[
S(F) = -\sum_{i=1}^{k} f_i \ln f_i
\]

A.1.7

The two constraints on the maximization are then, the \( \chi^2 \) constraint

\[
\Phi(F) = \sum_{i=1}^{k} \frac{(p_i - f_i)^2}{\sigma^2} - k = 0
\]

A.1.8

and the intensity constraint

\[
\Psi(F) = \sum_{i=1}^{k} f_i - 1 = 0
\]

A.1.9

The constrained maximum of \( S(F) \) is then obtained by solving eq. A.1.10 for each \( f_i \).
where \( \lambda_1 \) and \( \lambda_2 \) are Lagrange multipliers. Writing eq. A.1.10 explicitly for each \( f_i \) yields

\[
-1 - \ln(f_i) - \frac{2\lambda_1}{\sigma^2}(f_i - p_i) - \lambda_2 = 0
\]

A.1.11

For given values of \( \lambda_1 \) and \( \lambda_2 \) maximization of the constrained entropy is then achieved by finding the roots of eq. A.1.11. To find a solution which satisfies eq. A.1.8 and eq. A.1.9 it is necessary to find the values of \( \lambda_1 \) and \( \lambda_2 \). In our algorithm this is achieved by defining a function \( D(F) \)

\[
D(F) = [\left( \frac{\chi^2}{k} - 1 \right)^2 + (\sum_{i=1}^{k} f_i - 1)^2]^{1/2}
\]

A.1.12

\( D(F) \) may be regarded as a "distance" to solution. The values of \( \lambda_1 \) and \( \lambda_2 \) are then found using a downhill simplex search algorithm to find the minimum of \( D(F) \), which by definition will be at \( D(F) = 0 \). The algorithm will continue to search for the values of \( \lambda_1 \) and \( \lambda_2 \) until the constraints (equations A.1.8 and A.1.9) are satisfied. Solution of this problem requires the use of double precision calculation (16 bit math).

2.1.2 The \( E^2 \)-constrained Maximum Entropy Algorithm

Formally we define an array of residuals \( U = \{u_i: i=1, k\} \) such that

\[
u_i = \frac{1}{\sigma}(p_i - f_i)
\]

A.2.1

and a distribution for the noise, in our examples Gaussian noise, as \( Y = \{v_i: i=1, k\} \) where

\[
v_i = F^{-1}(i - 1/2)
\]

A.2.2

are the expected values of noise calculated from the inverse of the noise distribution, \( F \), defined.
\[ F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left( -\frac{y^2}{2} \right) dy \] \hspace{1cm} A.2.3

The error fitting constraint, \( E^2 \), is then defined as

\[ E^2 = \sum_{i=1}^{k} (u_{(i)} - v_i)^2 \] \hspace{1cm} A.2.4

where the subscripted parentheses indicate an ordering of the array \( U \) such that

\[ u_{(1)} < u_{(2)} < \ldots < u_{(k)} \] \hspace{1cm} A.2.5

The algorithm proceeds in a similar manner to that described for the \( \chi^2 \) constraint but \( \Phi(F) \) is now defined in terms of the error fitting constraint derived from eq. A.2.4 as

\[ \Phi(F) = E^2 - \ln(\ln(k)) = 0 \] \hspace{1cm} A.2.6

where \( \ln(\ln(k)) \) is the expected value of \( E^2 \). A model is found by using the Newton-Raphson method to find the root of the first differential of the constrained entropy equation. A downhill simplex search method is used to find the values of the Lagrange multipliers (equation A.1.10) that give a solution \( F \) which satisfies the two constraints (equations A.1.9 and A.2.6). A new distance function for the simplex to minimize must be defined as

\[ D(F) = \left[ \left( \frac{E^2}{\ln(\ln(k))} - 1 \right)^2 + \left( \sum_{i=1}^{k} f_i - 1 \right)^2 \right]^{\frac{1}{2}} \] \hspace{1cm} A.2.7

In order to assign the first values of \( u_i \) the algorithm requires the assignment of a starting model. This is usually taken to be the mean of the data set, the global maximum entropy model.
References


CHAPTER 3

A POSTERIORI DETERMINATION OF RELATIVE PROJECTION DIRECTIONS
OF ARBITRARILY ORIENTED MACROMOLECULES

Neil A. Farrow and F. Peter Ottensmeyer

In the generation of three dimensional structures from two dimensional projections a fundamental requirement is knowledge of the individual projection directions. In electron microscopy of single macromolecules the molecular projection directions are not known and must be determined a posteriori from the projection images themselves. The algorithm presented achieves such a determination using the central section theorem, geometrical techniques and quaternion mathematics. The quality of the solution is tested in relation to image noise, angular error in the input data, number of inter-compared projections used to generate common-axis data, and number of iterations. Correct determination of mutual alignments is achieved despite significant errors in the input data, indicating the method should be applicable to electron microscopy problems.

Accepted for publication Journal of the Optical Society of America, May 1992
3.1 Introduction

The reconstruction of a three dimensional (3D) density distribution of an object from a series
of its two dimensional (2D) density projection images requires prior knowledge of the 3D
orientation of these images with respect to each other. Frequently, as is the case in X-ray
tomography,1 multiple images of the same object are taken sequentially at different but known
angles. Similarly, in the electron microscope, the use of a microgoniometer support stage
permits the acquisition of a series of projection images at known orientations with respect to
the electron beam. Hoppe23 and co-workers have effectively exploited this technique, using
up to 20 different projections, to determine the structure of a number of molecules.44 The
combination of such tomographic series with the amenability of certain biological samples
to 2D crystallization6 7 has also provided high resolution 3D biological structures through the
exploitation of crystallographic redundancy.

Concerns about susceptibility of biological samples to radiation damage by the
imaging electron beam6 8 11 led to the desire to determine 3D structures from a minimum
number of exposures. In the special case that the molecule under investigation adopts a
preferred orientation with respect to the foil used to support it, the use of a single image of
the tilted sample allows 3D reconstruction using the random conical tilt technique.12 This
technique uses a second image only to determine the azimuthal angle of each image with
respect to an arbitrarily chosen reference orientation of the particles with respect to one
another. A number of molecules have been reconstructed using this method.13 14

This paper treats the general case for single images of a large number of identical
molecules which lie at random orientations to the electron beam. We present a scheme for
determining the relative angles between these orientations, which are calculated a posteriori
from the projections before a 3D reconstruction of the molecule is performed. The algorithm
described combines the central section theorem with quaternion mathematics. The sensitivity
of the alignments with respect to algorithm parameters and noisy input data is illustrated
using experimental simulations.
3.1.1 The Central Section Theorem

Virus structures possessing icosahedral symmetry were amongst some of the earliest biological structures to be reconstructed\textsuperscript{15} in 3D. The early workers\textsuperscript{16,17} recognized that the central section theorem may be employed to determine the relative projection directions of different views of the virus structures. The theorem states that, in Fourier space any two 2D Fourier transforms of corresponding 2D projections of a 3D density distribution will both be central sections of the 3D Fourier transform of the 3D density distribution, and thus they will share a common line in this 3D Fourier transform. Equivalently in real space this concept can be expressed by the statement that any two projections of a 3D density distribution will share a common line projection. By relating the positions of common line projections (common lines in Fourier space) to one another in a 3D coordinate system, the projection directions may be determined.

The central section theorem is pivotal to the method described in this chapter. We have applied it to the more general situation of asymmetrical objects.

3.1.2 Previous Work with Asymmetrical Objects

A number of authors\textsuperscript{18-21} have approached the determination of the relative angular alignment of projections with a variety of techniques. Most of these techniques have relied on the central section theorem. Both van Heel\textsuperscript{18} and Goncharov et al\textsuperscript{19} have formulated the problem in real space and Goncharov et al\textsuperscript{19} have also formulated it in Fourier space. An alternative solution proposed by Teague\textsuperscript{20} and implemented by Goncharov and Gelfand\textsuperscript{21} uses statistical methods to determine the relative orientation of a number of projections based on the moment information in those projections.

The method described below also uses the central section theorem to orient the projections. The procedure begins by using the geometrical approaches described by van Heel as angular reconstitution, and by Goncharov as the geometrical method. The work of these authors considers only a small number of projections, typically three. This paper describes the extension of these solution methods to a large number of projections. To accomplish this,
the central section theorem was combined with the use of quaternion mathematics. The input data for the technique described is information about the relative positions of common axes between different projections. For perfect data only two new common axes are necessary to align a subsequent projection to previously aligned projections. However, for imperfect data, knowledge of the positions of a larger number of common axes is advantageous. Since the algorithm is able to use all the common-axis information that is available to align a single projection, projections may be successfully aligned even when the input data has significant errors.

3.1.3 Experimental Constraints

From the central section theorem it is seen that any two 2D projections of the same 3D object will share a single common 1D line projection. The angles at which these line projections are generated specifies the position of the common axis between the projections. Functions referred to as sinograms may be employed to determine the position of the common line projection. A sinogram may be thought of as a collection of all the possible line projections (typically at 1° intervals) that may be derived from a projection image. The sinogram is thus a function of position (perpendicular to the projection direction) and the projection direction, usually expressed as an angle. To find the common line projection between two 2D projection images two such sinograms (one for each projection) must be calculated. The line projections contained in the two sinograms are then compared such that each line projection in one sinogram is compared to every line projection in the other. The line projections are compared by calculating the cross-correlation coefficient between them. The distribution of this coefficient may then be displayed as a two-dimensional function referred to as a sinogram correlation function. Examples of sinograms are shown in Fig. 1. (An example of a sinogram correlation function is shown in Chapter 5, Fig. 8.)

To estimate the level of the errors that are likely to be encountered in the positions of the common axes as a result of noise in electron microscopy data sets, the following simulation was performed. Twenty five pairs of projections of the 3D density distribution calculated from the X-ray diffraction structure of the Klenow fragment of DNA...
Figure 1. An example of the simulated projection images used to determine the error in common axis positions derived from the use of the sinogram correlation function. Projection images (A) to (F) illustrate the deterioration of image quality as noise is added to a simulated projection image of the Klenow Fragment of DNA Polymerase I generated from X-ray crystallography data. The fractional noise level in projection A: 0.00; B: 0.02; C: 0.05; D: 0.10; E: 0.20 and F: 0.50. Panel G illustrates the sinogram derived from the projection in panel A, panel H illustrates the sinogram generated from the image in panel F. The deterioration of the sinogram is clearly seen. (In displaying the sinograms the constant line projection arising from the uniform background intensity level in all projections has been subtracted.)

Polymerase I (Mol. Weight 68,000) were generated at known Euler angles. To determine the angular positions of the common axes the images were then used to generate a series of sinograms, and a sinogram correlation function was generated in the manner described above. The angular positions of the maxima in the sinogram correlation function were compared to those predicted from the known Euler angles. Subsequently, Gaussian noise was added pixel by pixel to the projection images and the effect on the accuracy of the angle determination from the sinogram correlation function was determined. The level of noise added to each projection is expressed as the ratio between the standard deviation of the noise added and the
Figure 2. The angular error between experimentally derived positions of common axes between pairs of projections and their theoretically predicted correct positions as a function of the fractional noise level introduced in simulated projections images. Each dot in the figure indicates the error for a particular pair of projections at a given noise level. The error bars indicate the standard error of the mean. The increase in the error of the determined common axis position with image noise is indicated by the slope of the line joining the means at each fractional noise level.

average intensity range for all the simulated images; thus a fractional noise level of 0.2 corresponds to the usual definition of a signal to noise ratio of 5. In our hands, digitized dark-field electron micrographs of unstained individual molecules typically have a fractional noise level of approximately 0.2. In Fig. 1 we illustrate the deterioration of the simulated image quality as the fractional noise level was increased from 0 to 0.5. We also show how the inclusion of noise in the projection images affects the structure of the sinograms generated from such projections.

In Figure 2 we have plotted the error between the experimentally determined position
of the common axis and the position calculated analytically from the known input orientations of a given pair of projections. As can be seen from Fig. 2 at a fractional noise level of 0.2 the average error in the position of the common axis is around 32°. Determination of the position of common axes using the sinogram correlation function is sensitive to noise in the projections because the function is usually relatively flat near its maxima. The angular errors indicated in Fig. 2 serve as reference when compared to the errors introduced into model data later in this chapter. However, it should be recognized that the simple Gaussian noise introduced into the projection images in the above study is only an approximation to the features that destroy signals in electron micrographs. The above noise model does not consider the much more complicated variations in the structure of interest that may occur as a result of specimen damage from the electron beam and the structure of the carbon support on which the molecule lies.

3.2 Coordinate Systems and Nomenclature

3.2.1 Euler Angles and Rotation Matrices

The 3D density distribution of an object will be denoted by \( P(x,y,z) \) where \( x, y \) and \( z \) represent three orthogonal axes which describe the "basis" coordinate system. Rotation of this coordinate system is described by a clockwise system of Euler angles as illustrated in Fig. 3. Euler angle \( \alpha \) denotes a rotation about the \( z \)-axis, \( \beta \) a rotation about the intermediate \( y \)-axis and \( \gamma \) a rotation about the new \( z' \)-axis. This rotation can be described by a matrix, \( R \), which is defined

\[
R = \begin{bmatrix}
\cos(\alpha)\cos(\beta)\cos(\gamma) - \sin(\alpha)\sin(\gamma) & -\cos(\alpha)\cos(\beta)\sin(\gamma) - \sin(\alpha)\cos(\gamma) & \cos(\alpha)\sin(\beta) \\
\sin(\alpha)\cos(\beta)\cos(\gamma) + \cos(\alpha)\sin(\gamma) & -\sin(\alpha)\cos(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma) & \sin(\alpha)\sin(\beta) \\
-\sin(\beta)\cos(\gamma) & \sin(\beta)\sin(\gamma) & \cos(\beta)
\end{bmatrix}
\]

The new frame of reference, denoted \( x', y' \) and \( z' \), after such a rotation will be described as the "projection" coordinate system. The 3D density distribution in this new system is described by \( P'(x',y',z') \). The 2D projection in the \( z' \) direction of the projection coordinate
The problem of determining the mutual orientation of planes is equivalent to that of determining the individual \( T \) matrices for each projection. The matrix that transforms the projection coordinate system of a projection \( a \) to the basis coordinate system is denoted \( T_a \).

### 3.2.2 Common Axes

The position of the common axis between two projections is determined by the use of Sinogram Correlation Functions as described by van Heel.\(^{18} \) The two angles that correspond to the maximum of the Sinogram Correlation Function define the position of the common axis in each of the projections. These angles express the position of the common axes relative to the \( x' \)-axes of their projection coordinate systems. The angular position of the common axis
between projections $a$ and $b$ in the coordinate system of projection $a$ is denoted $\theta_{ab}$; similarly $\theta_{ba}$ denotes the angular position of the same common axis in the system of projection $b$. The common axis may be written as a vector in the appropriate projection coordinate system. If the angle is measured in a counter-clockwise direction from the $x'$-axis then the common axis between projections $a$ and $b$ in the system of projection $a$ can be represented by the vector $c'_{ab}$ which is defined.

$$c'_{ab} = (\cos(\theta_{ab}), \sin(\theta_{ab}), 0)$$

The sinogram correlation function has a $180^\circ$ symmetry and it is possible either to chose vector $c'_{ab}$ to represent the common axis between projections $a$ and $b$ or to chose $-c'_{ab}$. In the following discussion the equalities will hold regardless of which of these two vectors is chosen so long as neither vector in a pair of common-axis vectors is reversed unilaterally. An incorrect initial choice of vector directions will eventually lead to the wrong-handedness of the 3D reconstruction. This ambiguity is intrinsic to the use of projections at unknown orientations.\textsuperscript{16}

The reorientation of the projection coordinate systems into the common basis coordinate system is accomplished by considering the positions of the common axes within the basis 3D system. The common axis $c'_{ab}$ will reorient itself under the correct transformation $T_a$ to the position $c_{ab}$. Correct solution of the mutual orientations of projection $a$ and $b$ should result in the following equation being true.

$$c'_{ab}T_a = c_{ba}T_b$$

In the basis coordinate system $c_{ab}$ should be equivalent to $c_{ba}$.

### 3.3 Orientation of Three Projections in a Single Coordinate System

Appendix 1 describes how three arbitrarily oriented planes may be aligned with respect to each other in a single coordinate system. The methods described are similar to those described by van Heel.\textsuperscript{18} To allow extension of the method to greater than three projections,
Figure 4. The alignment of projections 1, 2 and 3 determines the position of projection 4.

it was necessary to make some changes to the conventions used. Specifically, van Heel sets the common axis of projections 1 and 2 to be coincident with the z-axis of the basis coordinate system. In the following implementation one of the projection coordinate systems, namely that of projection 1, is defined to be coincident with the basis coordinate system. This change readily allows the incorporation of multiple projections with reference to the common basis coordinate system.

When the mutual orientations of the first three projections have been determined (as described by matrices $T_1$, $T_2$ and $T_3$) the addition of further projection planes to the basis coordinate system may be considered.

3.4 Addition of a Fourth Projection

The mutual orientations of three projections in a single coordinate system have been determined by satisfying the geometrical constraints on their mutual positions. The determination of the orientation of subsequent projections within the basis coordinate system is also possible using similar geometrical techniques. However, if the data contains errors, as will normally be the case in experimental applications, then such methods are likely to fail. Failure will occur because it is no longer possible to satisfy all the geometrical constraints that specify the positions of the common axes vectors for additional projections within the
same basis coordinate system. To illustrate this problem, consider the addition of a fourth projection to a basis coordinate system that contains three previously aligned projections. Figure 4 illustrates the data available to the next step of the orientation procedure. The first three Projections, denoted 1, 2 and 3, have been brought together in the basis coordinate system. If it is assumed that the common axes between these 3 projections and a fourth projection have been calculated then the positions of vectors \( c'_1 \), \( c'_2 \) and \( c'_3 \) in their respective projections coordinate systems are known. The transformations \( T_1, T_2 \) and \( T_3 \) will determine the positions of the vectors \( c_{14}, c_{24} \) and \( c_{34} \) respectively in the basis coordinate system (assuming that \( c_{ab} = c'_{ab}T_a \)). The corresponding vectors \( c'_{41}, c'_{42} \) and \( c'_{43} \) will also be known in the coordinate system of projection 4.

If the input data contains no errors and the orientations of the first three projections have been determined correctly then the vectors \( c_{14}, c_{24} \) and \( c_{34} \) will all lie in a single plane in the basis coordinate system. The angles between these vectors in the basis system should be equivalent to the angles between their counterpart vectors \( (c'_{41}, c'_{42} \) and \( c'_{43} \)) in the \( X'Y' \) plane of the coordinate system of projection 4. The introduction of errors in the determination of the position of the common axes in the projection systems will likely lead to \( c_{14}, c_{24} \) and \( c_{34} \) not being coplanar and the angles between the counterpart vectors not being equivalent in the two coordinate systems (i.e. the position of a general common axis \( c_{ab} \) is not identical to that of \( c'_{ab}T_a \) due to the data errors). It will then not be possible to determine a matrix \( T_4 \) which rotates vectors \( c'_{41}, c'_{42} \) and \( c'_{43} \) to exactly the positions \( c'_{14}T_1, c'_{24}T_2 \) and \( c'_{34}T_3 \) respectively. Thus we search for an optimum \( T_4 \) rotation matrix. The optimum \( T_4 \) is defined as the matrix that maximizes the sum of the scalar products between the pairs of vectors, i.e. the rotation matrix that maximizes the following expression

\[
\sum_{i=1}^{3} c'_{ai}T_i c'_{4i}T_4
\]

As discussed above, \( T_4 \) cannot be determined analytically using geometrical techniques due to the errors in the data. Maximization techniques such as the modified Powell Direction Set search using directions \( \alpha, \beta \) and \( \gamma \) fail as a result of becoming trapped in local maxima.

To solve the problem of the determination of \( T_4 \) the quaternion representation of
rotations described by Harauz\textsuperscript{24} and the solution techniques described by Horn\textsuperscript{22} were adopted. For a detailed description of the mathematics and conventions involved in quaternion mathematics the reader is referred to Horn. Appendix 2 describes the application of quaternions to the problem presented here, specifically the determination of $T_4$ which maximizes the expression in equation 5. The techniques described by Horn provide an extremely efficient and elegant mechanism for the determination of optimum rotation matrices under constraints similar to those described above.

3.5 Extension to a Large Number of Projections

As each projection is added to the basis coordinate system, more common-axis vector positions become known within this system. Thus for each new projection considered more vectors are available to determine the orientation of the subsequent projection using the quaternion method. The algorithm is designed to work through all the projections in turn, aligning each new projection using the common axes it has with the projections already aligned.

It is not necessary to determine all the common axes between all the projections. A minimum of two common axes are needed to determine the rotation matrix of each new projection. The advantage of determining more than two common axes arises when the determination of the position of these axes in the projection coordinate system is liable to be in error. It was assumed that the more common axes data available the more likely that the projection will be correctly oriented with respect to the other already determined projections. The validity of this assumption is discussed later in the context of the results shown in Fig. 8. It would seem logical to determine all the common axes between all the projections. However, a compromise is necessary, since, currently, generation and searching of a sinogram correlation function requires approximately 4 minutes CPU time on a VAX II computer. (Generation of a sinogram at $1^\circ$ intervals from a 64$^2$ pixel image requires 6 minutes CPU time on the same machine.)

The algorithm generates a scheme determining which pairs of projections will be compared to provide common-axis data. A variable referred to as the number of “check”
Figure 5. The matrix indicates a scheme generated for 16 projections with 3 check projections. A “1” in the matrix indicates the calculation of the position of the common axis $c_j$ between projections $i$ and $j$. Calculations are flagged only for the positions for which $j > i$. The existence of $c_j$ implies that the value of $c_j$ would also be known.

The algorithm proceeds by looking for three projections in this matrix for which common axes $c_ab$, $c_ac$ and $c_bc$ have been calculated; $a$ is normally set to 1 and $a < b < c$. When

<table>
<thead>
<tr>
<th>Projection $j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
three such projections are found, an attempt is made to align them using the geometrical method described in appendix 1. If this alignment fails (due to geometrical considerations) then $b$ and $c$ are increased until a successful alignment is achieved.

Following the determination of the mutual 3D orientation of the first three projections the algorithm proceeds to orient a fourth projection $c+1$, within the same basis coordinate system. The fourth projection is aligned using the available data relating to the common axes it shares with the previously aligned projections. Projection $c+1$ is aligned using the quaternion techniques. After each projection has been aligned this fact is recorded to indicate that common axes with that projection may be used to align other subsequent projections.

The process of aligning each projection by quaternion calculation of the rotation matrix is repeated until the final projection has been aligned. The algorithm then begins an iterative refinement procedure. Using the existing data, each projection is realigned with respect to the other aligned projections with which it shares common axes data. This iterative refinement process is implemented to improve the quality of the solution by gradually removing biases in the data that may have been imposed by the initial order of the alignment and selection of the first three projections. The effect of changing the number of refinement iterations is discussed in section 3.8. Once the values of $\theta_y$ and $\theta_Z$ have been determined, the algorithm requires only approximately 50 seconds CPU time on a \muVAX II computer to perform the alignment calculations for 32 projections using 5 check projections and 64 iterations.

3.6. Measures of Algorithm Performance

Three parameters have been used to assess the performance of the algorithm; self-consistency error (S.C.E.), coordinate system error (C.S.E.) and residual system error (R.S.E.). The algorithm uses the property described in equation 4 to determine individual projection directions. However, as discussed earlier in section 3.4, erroneous positions of $\theta_y$ and $\theta_Z$ will result in equation 4 not necessarily being satisfied for all common axes. The S.C.E. characterizes the degree to which the algorithm has been able to determine the mutual orientations of the individual projections in a manner consistent with the input data whilst
attempting to satisfy equation 4 for all common axes. The S.C.E. is a measure of the average angular discrepancy between the positions of $c'_iT_i$ and $c'_jT_j$, both vectors in the basis coordinate system. The S.C.E. is written explicitly as

$$S.C.E. = \frac{1}{N_c} \sum_i \sum_j \arccos( c'_iT_i,c'_jT_j)$$

where $N_c$ is the total number of common axes considered in the data set. Thus, if the input data were to contain no errors and the projection directions were correctly determined then the value of S.C.E. would be zero, indicating equation 4 holds exactly for all common axes.

Whilst the S.C.E. provides a measure of the degree to which the algorithm has satisfied internal alignment constraints caution has to be exercised in interpreting the overall accuracy of the alignment of the projections from the S.C.E. The S.C.E. is influenced not only by the consistency of the alignment but also by the consistency of the input data. For instance, if the projection directions were determined correctly but the initial angular positions of the common axes in the projection coordinate systems were in error, either due to experimental causes or as in the below simulations because of the introduction of errors, then S.C.E. will to some extent reflect the errors in the input data.

In the simulations described below, errors were introduced in the angular positions of the common axes by the addition of an error angle uniformly distributed over $+/\Delta$. A simple probabilistic geometrical analysis of the angular errors reveals that if the projection directions are assigned correctly then the expected value of the S.C.E. will be $0.83\Delta$ (equivalent to the average angular error in the basis system between the pairs of vectors representing a single common axis in the input data.) The S.C.E. may be regarded as an internal quality measure for the algorithm.

As model data were used to study the performance of the method the correct orientations of the projections were known. This information was used in the other measures of algorithm performance. After the projection directions had been experimentally determined the positions of the projection coordinate systems were compared to those used to create the original input data.

Direct comparison of the values of the Euler angles does not provide a useful measure of the accuracy of the determination of projection directions; a combination of vastly different
Euler angles can yield similar projection directions. A reliable measure of the accuracy of assignment of projection directions may be obtained by comparing the position of the vectors representing the x and z-axes of each projection coordinate system in the basis coordinate system using both the experimentally determined rotation matrices $T_i$ and the matrices that were used to generate the original data set, which we denote $U_i$. The coordinate system error, C.S.E., then measures the average angular discrepancy between these pairs of axes, explicitly,

$$\text{C.S.E.} = \frac{1}{2N_p} \sum_{i=1}^{N_p} \left( \arccos(x_{T_i} x_{U_i}) + \arccos(z_{T_i} z_{U_i}) \right)$$

where $N_p$ is the total number of projections under consideration.

The C.S.E. can be regarded as an absolute error between the orientations of the coordinate systems used to generate the original data and those determined from that data. However, it is important only that the algorithm determines the relative projection directions correctly. A general rotation of all the experimentally determined projection directions would be seen as an error in the C.S.E. To separate the effect of such a rotation we calculate a measure referred to as the residual system error, R.S.E. This error is defined

$$\text{R.S.E.} = \frac{1}{2N_p} \sum_{i=1}^{N_p} \left( \arccos(x_{T \cdot i} x_{U_i}) + \arccos(z_{T \cdot i} z_{U_i}) \right)$$

The measure is essentially similar to that described by the C.S.E. but an optimal general rigid-body rotation is applied to all the projection coordinate systems described by the experimentally determined angles. The matrix describing this general rotation, $T_g$, may be determined by an extension of the Quaternion mathematics described by Horn. The matrix maximizes the expression

$$\sum_{i=1}^{N_p} \left( x_{T \cdot g_i} x_{U_i} + z_{T \cdot g_i} z_{U_i} \right)$$

The R.S.E. can thus be regarded as a C.S.E. type measure corrected by a general rotation.

87
3.7 Simulations

A number of simulations were performed to determine the efficacy of the methods described above in recovering projection directions \textit{a posteriori} from common-axis data. In all of the following simulations the input data for the algorithm consisted of the angular positions $\theta_j$ and $\theta_k$ of a common axis between two projections. The projections directions (to be determined \textit{a posteriori} by the algorithm) were assigned using a random distribution of Euler angles. The common axis between any two projections is equivalent to the vector representing the intersection of the $X'Y'$ planes of their respective coordinate systems. The angular position of the common axis is thus equal to the angle that the intersection vector makes with the $x'$-axis in each projection system. Thus the input data in the simulations effectively reproduced data that would be available after the maxima of the Sinogram Correlation Functions had been located. To allow for a more realistic simulation of experimental data an error angle was added to the common-axis positions. In the following experiments an error of $\Delta^\circ$ indicates that an error angle selected from a uniform distribution between $\Delta^\circ$ and $-\Delta^\circ$ was added to each common-axis position.

The size of this error was one of the variables declared at the beginning of each simulation. The other variables were the total number of projections, the number of check projections and the number of alignment iterations through all the projections that would be performed.

3.8 Results and Discussion

Figure 6 shows the relationship between changing the number of iterations used to determine the projection directions and the quality of the solution achieved. The problem given to the algorithm was to determine the projection directions of 32 projections using 5 check projections when the common-axis data had errors of $5^\circ$. The number of iterations ranged from 2 to 64. Figure 6 indicates that the \textit{S.C.E.} does not decrease greatly as the number of iterations is increased. The \textit{C.S.E.} does decrease as the number of iterations is increased while the \textit{R.S.E.} follows a pattern more similar to that of the \textit{S.C.E.} The \textit{R.S.E.} value stabilizes after
Figure 6. The effect of the number of iterations used to find an alignment solution on the quality of that solution. The quality of solution is expressed by three measures, the self-consistency error (S.C.E.), the coordinate similarity error (C.S.E.) and the residual system error (R.S.E.).

around 30 iterations. The form of the three curves may be explained as follows. The alignment after the second pass through all the available common-axis data is fairly close to the best configuration possible. This is reflected in the small changes in the S.C.E. and R.S.E. with further iterations. The coordinate system error decreases with increasing numbers of iterations as errors introduced into the total system from positional errors in the first few projections are removed. As these errors are removed using information available from other projections the C.S.E. decreases and tends to the value of the R.S.E. The values of C.S.E. and R.S.E. will not necessarily become equal; the algorithm does not have the information that is available when calculating the R.S.E. (i.e. the true projection directions). When 64 iterations were used, the corresponding C.S.E. and R.S.E. values of the solution were 1.98°
Figure 7. The effect of the error in the common axes positions used as input data to the alignment procedure on the quality of the solution. The quality of solution is expressed by three measures, the self-consistency error (S.C.E.), the coordinate similarity error (C.S.E) and the residual system error (R.S.E). The error bars indicate the standard deviation of the result in 10 experiments. Calculations were carried out every 5°: the error bars have been offset slightly for clarity.

and 1.41° respectively. For the same data set, 300 iterations result in values of C.S.E. and R.S.E. of 1.83° and 1.41° respectively.

The final self-consistency error can be compared to the expected self-consistency error for a given angular error assuming correct determination of random projection directions. The expected S.C.E. is 0.83 times the error introduced in the positions of the common axes. Thus for an error of 5° the value of S.C.E. at the correct alignment is expected to be 4.15°. The final solution to the problem in Fig. 6 has an S.C.E. of 3.4° indicating that the orientation of
the planes have been determined with a slight error.

We interpret the finding that the experimental S.C.E. is lower than the expected S.C.E. in the following manner. The alignment of each projection is facilitated by finding the matrix $T$, which maximizes the expression in equation 5. The algorithm makes no allowance for errors in the input data. If the position of a projection is defined by only a few common axes, as in the above experiment, then a set of projections can often be found whose alignment is internally consistent with an S.C.E. lower than for the optimum set of projection directions (see discussion of Fig. 8).

Figure 7 indicates how the quality of the solution is affected by the size of the errors in the common-axis data. All experiments were carried out using 32 projections, 5 check projections and 64 iterations. The error in the common-axis data was increased from $0^\circ$ to $40^\circ$ and at each error level, 10 different alignment experiments were performed. For $0^\circ$ angular error the algorithm determines the relative angular orientations exactly. As expected, the S.C.E. increases as the error in the input data increases. The slope of the S.C.E. line indicates that the S.C.E. is on average 0.628 of the error added to the common-axis data. Again this error is somewhat smaller than that expected from the original data for the reasons given above.

The C.S.E. and R.S.E. plotted in Fig. 7 show that both errors increase linearly as the noise in the data increases with slopes of 0.45 and 0.33 respectively. However the value of R.S.E. even at relatively high levels indicates that the orientation of the projections is still being determined effectively. Figure 2 indicates that at a fractional noise level of 0.2 (S.N.R. of 5) the mean error in the determination of the position of the common axis is $32^\circ$. While we do not model the error distribution exactly in these simulations by the addition of a uniformly distributed error angle, when this error angle is $32^\circ$ the method was able to determine projection directions to within a R.S.E. of $10^\circ$.

Figure 8 illustrates the effect of increasing the number of check projections on the quality of the solution. In each experiment, 32 projection directions were to be determined using 64 iterations from data with an error of $5^\circ$. The number of check projections was increased from 3 to 25 and at each check level, 5 different alignment experiments were performed. From the figure it can be seen that the self-consistency error increases very
Figure 8. The effect of the number of check projections used in the alignment procedure on the quality of the solution. The quality of solution is expressed by three measures, the self-consistency error (S.C.E.), the coordinate similarity error (C.S.E.) and the residual system error (R.S.E.). The error bars indicate the standard deviation of the result in 5 experiments. The error bars have been offset from the actual calculation points slightly for clarity. See main text for description of the significance of the two marked outlying points.

slightly as the number of check projections increases. This result appears counter-intuitive but can be readily explained. Firstly it should be remembered that the self-consistency is non-zero and ideally equal to $0.83\Delta^\circ$ when a correct alignment is determined if the common-axis data had an error added of $\Delta^\circ$. However, the use of quaternions in maximizing the expression in equation 5 will minimize the S.C.E. in general to a level below this value. As the number of projections increases the freedom of the algorithm to maximize the expression in equation 5 decreases, since more constraints are put on the alignment of each projection. The S.C.E.
should then be expected to rise to the value predicted of 4.1° for an error of 5°.

As expected the C.S.E. and the R.S.E. both improve as the number of check projections increases. This is interpreted as being due to the position of each projection becoming better defined in the basis coordinate system as the number of vectors in each projection plane is increased.

One particular solution has been plotted separately in Fig. 8. This solution occurred when 5 check projections were used. The quality of this solution was relatively poor, S.C.E., C.S.E. and R.S.E. having values of 5.1°, 38.1° and 3.9° respectively. When the solution was examined to determine why such large errors may have occurred, it was discovered that the particular orientation of the first three planes had led to large errors being introduced into the subsequent alignment.

In the simulation experiments described above, the checking of the quality of the alignment solution was possible because the original projection directions were known. When implementing the method using real projection data, certain other precautions will be necessary to ascertain that the alignments determined are reproducible. The method occasionally produced alignments with a high error despite low error levels in the input data. These erroneous alignments were usually due to unfavourable geometries of starting projections. When determining common axes experimentally, if two projection directions are nearly parallel, then the position of the common axes is more likely to be determined with a large angular error. To overcome these problems it may be necessary to implement the algorithm starting with a number of different three projection sets to compare the relative positions of the aligned projections. In this manner it is believed that it will be easy to detect and reject incorrect alignments, as demonstrated by the observation of the outlying points in Fig. 8.

The success of the algorithm in determining the relative orientations of the projections has been quantified by the use of three novel quality measures (S.C.E., C.S.E. and R.S.E.). In practical situations only one of these measures, the S.C.E., would be available to characterize the success of the algorithm. We have not attempted to determine the relationship between the three measures, and some quantification of the resolution of a reconstructed structure due to the other variables (specimen dependency, distribution of projection
directions, reconstruction techniques employed etc.). Instead, the three measures we have used, permit some intuitive understanding of the angular errors that result from the alignment of the projections.

Another measure that might be applied to characterize the quality of experimental data sets could be similar to the comparison of the common lines of virus particles used by Crowther et al.\textsuperscript{15} The sum of residuals between the common lines was employed to determine the degree of preservation of symmetry to a given resolution limit. Similar measures may be implemented using the above data simply by taking the Fourier transform of the line projections corresponding to the maxima of the sinogram correlation function.

### 3.9 Conclusion

The principles of the central section theorem have been combined with geometrical techniques and the use of quaternion mathematics to permit the determination of a large number of projection directions \textit{a posteriori}. This procedure permits the alignment of projections in which there may be significant amounts of noise (S.N.R of 5). In the process all geometrical conditions are not necessarily satisfied exactly, allowing the algorithm to cope with errors in the data. The algorithm determines the best alignment for each projection by optimizing the match to predetermined geometrical constraints that in this work ranged from a minimum of 2 to 27 but are not bounded by this limit. Quaternion mathematics enables the determination of this optimum alignment. The quality of the alignments determined by the algorithm have been described as a function of the angular error in the common-axis data, of the excess number of projections used above the required minimum and of the number of iterations. For noisy input data it has been shown that increasing the number of inter-comparisons and the number of iterations gives rise to a corresponding increase in the accuracy of the final orientation determination. For every set of random starting orientations the algorithm determined the relative angles exactly (no noise), and was also sufficiently robust to determine the alignment to within the random errors introduced into the data.
3.1 Alignment of Three Projections

The conventions that describe the positions of the common axes in the three projection coordinate systems are shown in Fig. A.1. These three planes will be aligned with respect to each other, in the basis coordinate system, as illustrated in Fig. A.2.

As the coordinate system of projection 1 (Fig. A.2) is adopted as the basis coordinate system for the alignment then by definition the vectors $c_{12}$ and $c_{13}$ are expressed in vector notation as

$$c_{12} = (\cos(dA), \sin(dA), 0) \quad A.1.1$$

$$c_{13} = (\cos(\beta), \sin(\beta), 0) \quad A.1.2$$

The position of any general unit vector, $\mathbf{v}$, in the basis coordinate system (Fig. A.2) may be expressed in terms of two angles, $\alpha$ and $\beta$.

$$\mathbf{v} = (\sin(\beta)\cos(\alpha), \sin(\beta)\sin(\alpha), \cos(\beta)) \quad A.1.3$$

This vector is equivalent to the position of the z-axis of the basis coordinate system following
a rotation of $\alpha$, $\beta$, $\gamma = 0$ under the general matrix $T$. The common axis $c_23$ is represented in vector notation by

$$c_{23} = (\sin(\beta_{23})\cos(\alpha_{23}), \sin(\beta_{23})\sin(\alpha_{23}), \cos(\beta_{23}))$$

A.1.4

where $\alpha_{23}$ and $\beta_{23}$ are to be determined. The values of $\alpha_{23}$ and $\beta_{23}$ can be determined in the following manner. We know that in the basis coordinate system in which $c_{ij} = e_j$ that

$$c_{12}c_{23} = \cos(B+dB)$$

A.1.5

$$c_{23}c_{13} = \cos(C+dC)$$

A.1.6

Writing out the left side of these equations explicitly, it can be shown that

$$\tan(\alpha_{23}) = \frac{\cos(C+dC)\cos(dA) - \cos(B+dB)\cos(A)}{\cos(C+dC)\sin(dA) + \cos(B+dB)\sin(A)}$$

A.1.7

and
\[
\sin(\beta_{23}) = \frac{\cos(B + dB)}{\cos(\alpha_{23}) \cos(\alpha_{23}) - \sin(\alpha_{23}) \sin(\alpha_{23})}
\]

There are obviously two roots to each of these equations. Physically these roots represent the difference between solving the problem so that \(e_{23}\) has a positive or negative \(z\) component. Choice of which pair of roots is used for the subsequent calculations is made arbitrarily.

The \(z'\)-axis of the coordinate system of projection 2, \(z'_2\), can be represented in a similar manner to \(e_{23}\), thus

\[
z'_2 = (\sin(\beta_2)\cos(\alpha_2), \sin(\beta_2)\sin(\alpha_2), \cos(\beta_2))
\]

Angles \(\alpha_2\) and \(\beta_2\) can be determined by using the relationship between the \(z'_2\)-axis and the vector product of \(e_{23}\) and \(e_{12}\)

\[
z'_2 \sin(B + dB) = e_{23} \times e_{12}
\]

 explicitly

\[
(\sin(\beta_2)\cos(\alpha_2), \sin(\beta_2)\sin(\alpha_2), \cos(\beta_2)) \sin(B + dB) = \\
(\sin(\beta_{23})\cos(\alpha_{23}), \sin(\beta_{23})\sin(\alpha_{23}), \cos(\beta_{23})) \times (\cos(dA), -\sin(dA), 0)
\]

Equating the \(x\) and \(y\) components of the resultant vector

\[
\frac{\sin(dA)}{\cos(dA)} = \frac{\cos(\alpha_2)}{\sin(\alpha_2)}
\]

which implies \(\alpha_2 = 90^\circ - dA\) or \(\alpha_2 = 270^\circ - dA\). Angle \(\beta_2\) can similarly be recovered to give

\[
\sin(\beta_2) = \frac{\sin(dA)\cos(\beta_{23})}{\sin(B + dB)\cos(\alpha_2)}
\]

which again has two roots.

Using a similar representation for \(z'_3\) and the vector product of \(e_{31}\) and \(e_{23}\), it can be determined that \(\alpha_3 = 90^\circ + A\) or \(\alpha_3 = 270^\circ + A\). \(\beta_3\) is defined by the expression
The values of $\gamma_2$ and $\gamma_3$ may be determined by considering the rotation required to bring $e'_2$ and $e'_3$ to their correct positions in the basis system. It can be determined that $\gamma_2 = -(90^\circ + B)$ and $\gamma_3 = -(90^\circ - C)$.

It should be noted that if the common axes positions are in error then it is possible that a solution may not be found for the three projections. This would be the case if $(A+dA) > ((B+dB) + (C+dC))$.

It is necessary in the algorithm to determine which roots of the equations for $\alpha$, $\beta$, and $\gamma$ are to be used. This is done by simply checking that the positions of all the rotated common-axis vectors are consistent with one another in the basis coordinate system.

3.2 Determination of Rotation Matrices $T$ by Quaternion Methods

3.2.1 Properties of Quaternions

A quaternion can be thought of as a complex number with a single real component $q_0$ and three imaginary components $q_x$, $q_y$, and $q_z$. This quaternion is written symbolically as $q$. The three imaginary components are regarded as being in three orthogonal directions denoted $i$, $j$, and $k$. Thus

$$q = q_0 + iq_x + jq_y + kq_z$$

A scalar can be represented as a purely real quaternion i.e. $q_x = q_y = q_z = 0$ and a vector can be represented as a purely imaginary quaternion i.e. $q_0 = 0$. To discuss how the quaternions may be used to determine the optimum rotation of coordinate systems it is necessary to introduce a few concepts of the mathematics of quaternions.

i) A quaternion $q$ has a complex conjugate $q^*$ such that

$$q^* = q_0 - iq_x - jq_y - kq_z$$
ii) The products of the components of quaternions multiply as

\[ ii=jj=kk=-1; \quad ij=ik; \quad ji=k; \quad ji=-k; \quad kj=-i; \quad ik=-j \]

A.2.3

iii) The arithmetic product of two quaternions is given by

\[ pq = (p_0 q_0 - p_x q_x - p_y q_y - p_z q_z) + i(p_0 q_x + p_x q_0 + p_y q_z - p_z q_y) + j(p_0 q_y - p_y q_0 + p_x q_z + p_z q_x) + k(p_0 q_z + p_z q_0 - p_x q_y + p_y q_x) \]

A.2.4

From this definition it is apparent that multiplication of quaternions will not be commutative.

iv) Quaternion multiplication can be represented by multiplication of an equivalent 4x4 matrix. Thus

\[
\begin{bmatrix}
    p_0 & -p_x & -p_y & -p_z \\
    p_x & p_0 & -p_z & p_y \\
    p_y & p_z & p_0 & -p_x \\
    p_z & -p_y & p_x & p_0 \\
\end{bmatrix}
\begin{bmatrix}
    q_0 \\
    q_x \\
    q_y \\
    q_z \\
\end{bmatrix} = pq
\]

A.2.5

This 4x4 matrix is denoted \( P \). (The bold typeface is used to indicate a matrix representation of a quaternion). There exists a similar matrix \( \tilde{P} \) which is defined

\[
\tilde{P} =
\begin{bmatrix}
    p_0 & p_z & -p_y & -p_x \\
    p_x & p_0 & p_z & -p_y \\
    p_y & -p_z & p_0 & p_x \\
    -p_x & -p_y & -p_z & p_0 \\
\end{bmatrix}
\]

A.2.6

So products of quaternions can be represented in a vector matrix notation as

\[ pq = \tilde{P} q \]

A.2.7

and

99
In conventional vector notation, the rotation of a vector \( p' \) to \( p \) under a rotation described by matrix \( M \) can be written

\[
p = M p'
\]

In quaternion terms the above equation can be regarded as describing the mapping of one purely imaginary quaternion to another. We can represent the rotation using another unit quaternion \( m \). Horn shows that in order to preserve the imaginary nature of the quaternions representing the vectors it is necessary to write the transformation as

\[
p = m \bar{p}' m^*
\]

which may be shown to be equivalent to

\[
p = \tilde{m}^T M p'
\]

The matrix \( \tilde{M}^T M \) may be written explicitly as

\[
\begin{bmatrix}
    m \cdot \bar{m} & 0 & 0 & 0 \\
    0 & (m_x^2 + m_y^2 - m_z^2) & 2(m_x m_y - m_z m_w) & 2(m_x m_z + m_y m_w) \\
    0 & 2(m_y m_z + m_x m_w) & (m_y^2 + m_z^2 + m_x^2 - m_w^2) & 2(m_y m_z - m_x m_w) \\
    0 & 2(m_z m_x - m_y m_w) & 2(m_z m_y + m_x m_w) & (m_z^2 + m_x^2 + m_y^2 - m_w^2)
\end{bmatrix}
\]

Horn shows that the lower 3x3 submatrix of the above is exactly equivalent to the general matrix \( M \) of equation A.2.9 above.

3.A.2.2 Application to the problem of determining matrix \( T_A \)

Expression 5 in the main text describes the function that is to be maximized to allow the determination of the optimum \( T_A \). To simplify the notation we can rewrite this expression as
assuming that the true position of the common axes with projection \( c_{i*} \) are accurately specified by \( Tc'_{i*} \). Now in quaternion notation we may represent the common-axis vectors in the basis coordinate system, \( c_{i*} \), by \( \epsilon_{i*} \). The common axes in the projection coordinate systems \( c'_{i*} \) may be represented by the quaternions \( \epsilon'_{i*} \). The rotation matrix \( T_{i} \) will be represented by the quaternion \( \epsilon_{i} \). The expression equivalent to that of A.2.13 may now be written as

\[
\sum_{i=1}^{3} \epsilon'_{i*}(i_{i}^{*} c''_{i*} * i_{i}^{*})
\]  

A.2.14

Following the arguments of Horn it can be shown that this expression may be rewritten

\[
i_{i}^{*} \left( \sum_{i=1}^{3} \overline{C}'_{i*} c_{i*} i_{i}^{*} \right)
\]

A.2.15

in which the quaternions \( \epsilon'_{i*} \) and \( \epsilon_{i*} \) are represented by their equivalent 4x4 matrices. The summation now forms a new 4x4 matrix, the components of which are generated from the sums of products of the quaternions \( \epsilon'_{i*} \) and \( \epsilon_{i*} \) or equivalently the terms of the vectors \( c'_{i*} \) and \( c_{i*} \). Horn shows that the quaternion \( i_{i}^{*} \) that maximizes the above expression is given by the eigenvector associated with the largest positive eigenvalue of the 4x4 matrix that is derived from the sum in A.2.15. Having determined this quaternion using standard eigenvector computation routines, the matrix equivalent to expression A.2.12 can be generated. Using the equivalence between the lower right 3x3 matrix and the rotation matrix \( T \) we can determine the values of \( \alpha \), \( \beta \) and \( \gamma \) that define the rotation.
References


Determination of the relative orientation of 2D projection images of a 3D density distribution is required to permit reconstruction of the 3D distribution from the images. Using a technique that combines the common axis method and quaternion mathematics the projection directions may be determined \textit{a posteriori}. In this chapter the method is applied to simulated images of the Klenow fragment of DNA Polymerase I. The simulated images are subjected to the image processing procedures that would normally be applied to electron microscope images to improve their quality, namely 2D registration procedures and multivariate statistical classification techniques. Results are presented and a modification of the original algorithm is described. The alignment procedure is able to determine the relative orientations of the projections correctly even when the images have an initial signal to noise ratio of 3.1 and the images forming a single class of views are perturbed slightly.

A version of this chapter will be submitted to \textit{Ultramicroscopy} for publication.
4.1 Introduction

Electron microscopy has been used to determine the three-dimensional (3D) structures of a large number of biological macromolecules and macromolecular complexes.\textsuperscript{1-16} Reconstruction of a 3D molecular density distribution from its two-dimensional (2D) projection images, its electron micrographs, has required knowledge of the relative 3D orientation of different projections. Several methods have been proposed for the \textit{a posteriori} determination of projection directions of non-symmetrical objects which assume random orientations with respect to the support film or may be embedded at random orientations from the use of vitreous-ice embedding techniques.\textsuperscript{17-21} A number of these techniques\textsuperscript{17-19} employ the principles of the common axis theorem which are discussed in section 1.2 below. Others use the methods of moments to determine the relative orientations of projections.\textsuperscript{20,21} In the previous chapter an algorithm was described that extended the common-axis-based techniques to incorporate a large number of projections.\textsuperscript{22} The use of a large number of projections is important, since the spatial resolution of the 3D structure is related to the number of different projections available for reconstruction both in terms of angular distribution and in terms of noise.

The results in Chapter 3 demonstrated that the algorithm was able to determine projection directions when the input data was in the form of angular positions of common axes between hypothetical projection images. It was demonstrated that the algorithm was capable of determining projection directions accurately despite large errors in this type of input data, an \textit{R.S.E.} of just 12° was obtained when the input data had errors of up to ±40°.

In this chapter simulated data will again be employed to further study the application of the alignment technique. The major difference between this study and the one described previously is that here the input data is in the form of simulated projection images. In the previous study the positions of the common axes between images were determined analytically. In the work described in this chapter the positions of the common axes are determined directly from the images using the sinogram correlation functions described by van Heel.\textsuperscript{19} Thus the simulations described in this chapter form an important bridge between the theoretical simulations and results of the previous chapter and the practical application.
of the method to images from the electron microscope which will be described in Chapter 5. Because images are used as the initial data in the following simulations these studies extend the results of previous simulations to include the effects of errors that are likely to arise from the imaging and image processing techniques, as applied to dark-field electron micrographs.

A typical electron micrograph is derived as a compromise between the number of electrons used to reduce the statistical noise in the image and the damage to the sample caused by the imaging electrons. As a result, the images are usually taken with the lowest practicable dose commensurate with the desired spatial resolution. Another serious degradation of dark-field images arises from a noise source which is stationary in time, the structure of the support film, usually an ultra-thin carbon film upon which the molecule of interest lies. Local variations in the support film thickness will manifest themselves within the images as variations in image background intensity. One of the most common techniques employed to remove such variations from images is to identify classes of similar images and to average them. However, if the structures of interest in the different images to be averaged are not sufficiently identical, usually due to slightly differing projection directions, distortion of the molecule or different degrees of flexing between molecules, then the summation process itself will lead to a drop in the resolution of the summed images.

The major image processing steps that are likely to be employed between the collection of images and the eventual generation of a 3D structure can be separated into 3 distinct stages: firstly, the rotational alignment of images prior to an automatic classification procedure; secondly, the classification procedure itself; thirdly, the determination of the mutual alignment of the projections with respect to one another. Procedures for the classification of images have been discussed widely elsewhere. The preliminary rotational alignment and final projection direction determination steps are discussed more fully below.

In order to determine the behaviour of the algorithm under precisely known adverse but realistic conditions simulated dark-field electron micrographs were calculated as projections from the X-ray crystallographic structure of the Klenow fragment of DNA polymerase I, with defined signal to noise ratios and angular errors (see below).
4.1.1 Reference-Free Alignment Techniques

Automatic image classification algorithms\(^{26,27}\) with the exception of that proposed by Schatz and van Heel\(^{29}\) assume that the input data set contains images that are consistently aligned, such that any two similar views of a molecule would be in the same orientation with respect to the image coordinate system. Such an alignment is necessary to ensure that similar but initially rotated images are recognized and classified as having resulted from similar projection directions.

Rather than classifying the images directly, Schatz and van Heel propose that the classification be based on the double-autocorrelation function of the images which will be rotationally and translationally invariant. Despite a number of advantages in this approach, we chose not to adopt this classification technique for two reasons. Firstly, the phase information in the Fourier transform of the image is not retained with this procedure i.e. similarity of images is based only on the intensities of the spatial frequencies in the image and therefore, all the information available in the image is not used. Secondly, although double-autocorrelation functions can be used to classify images there remains the need to rotate and translate the original images prior to their summation within classes.

The most widely implemented alignment procedure is to align the images with respect to a reference image using a rotational cross-correlation technique\(^{30}\). A variation of this technique uses multiple reference images and aligns each image with respect to each of the references individually\(^{31}\). Implicit in this technique is the need for high quality reference images for the alignment. The multi-reference techniques were developed for applications in which the particles under study were observed to occupy only a limited number of orientations relative to the support film. When the molecules exhibit a quasi-continuous distribution of orientations, selection of high quality reference images becomes more problematic.

In this paper we employ a reference-free alignment technique. Although the principles of the technique were originally suggested to us by Penczek the procedure was developed independently by us and as a result varies slightly in implementation from that recently described by Penczek and Frank\(^{32}\). The most significant difference between our
implementation, which was developed using routines available in the IMAGIC image processing framework, and that described by Penczek and Frank is that in our algorithm the rotational and translational alignments of each image are combined. Our implementation of the reference-free procedure is as follows. From a large data set of images of molecules at unknown orientations the first two images in the set are aligned translationally and rotationally with one another. The second image is then summed with the first to form a composite image. A third image is then aligned with respect to this composite image and added to the composite. This process is repeated for each of the images in the data set. After the first alignment of all the images, each image is removed in turn from the composite and realigned with respect to the remainder of the composite image. This process is repeated for all the images in the data set. The realignment procedure may be repeated as many times as required for the alignment to stabilize. We have demonstrated, (unpublished results), that this technique gives rotational alignments that stabilize after just a few iterations (=5) and that the relative rotations of images are not influenced by the image used to start the alignment procedure.

4.1.2 Projection Direction Determination

A number of authors\textsuperscript{17-19} have proposed techniques using the principles of the common axis theorem and geometrical approaches to align a limited number of low-noise views of asymmetrical objects. We have extended the use of common axes as a tool for alignment determination to encompass an arbitrary number of projections images with low or high noise levels.

The alignment technique used here by us to determine the relative 3D orientation of projection images has been fully described elsewhere.\textsuperscript{22} A brief description of the procedure is given below. The method relies on the principles set out by the common axis theorem which were applied to the orientation of views of symmetrical virus structures.\textsuperscript{12} The principle itself may be stated in the following manner, any two 2D projections of a single 3D density distribution will share a common line projection. Determination of the position of this common line projection allows the relative orientation of two projections to be determined.
to within a single rotation angle. Comparison of the line projections of three 2D projections allows complete determination of the mutual orientation of the three planes (except for a degree of handedness).

The position of a pair of common axes between two different projection images of the same 3D object are determined using the sinogram correlation functions. Given common axis pairs between many such projection images, the algorithm first aligns three randomly selected projections in a single common 3D coordinate system using geometrical principles and constraints similar to those described by van Heel. Some changes in definitions were required to allow extension to a large number of projections. Once the first three projections have been aligned, the algorithm proceeds in a slightly different manner. The next projection and all subsequent ones must possess common axes between them and the projections for which 3D orientations have already been determined. In theory only two common axes between any current single projection and previous ones are needed to define its position. Due to noise considerations, however, common axes with a greater number of projections, typically five, are calculated. The orientation of the current projection image is determined by finding the Euler angles corresponding to the rotation of the image that minimizes the angular discrepancies between the already established positions of the common axes and the positions of the common axes associated with the projection under rotation. The optimum rotation is determined by the use of quaternion mathematics.

Once all projections have been aligned with respect to one another, each projection is removed in turn and realigned with respect to all the other projections with which it shares a calculated common axes. In this iterative manner the consistency of the alignments is improved and any biases introduced by errors in the first three projections used for the initial geometrical alignment will be reduced. For the remainder of this paper the above Iterative Quaternion-based Alignment Determination techniques will be referred to as IQAD.

4.1.3 Assessment of Algorithm Performance

For perfect data the pair of common axes between any two projections will be rotated to coincident positions in the coordinate frame that is fixed by the first set of three projections.
However, for a projection that contains noisy data or an angular error, a compromise is made by determining the rotation which is the best fit among the common axes between it and the other projections (see Appendix). As a result any corresponding pair of axes from two projections will not in general be coincident after such an optimum rotation. The average angular difference from coincidence for all such pairs has been used here as a measure of the success of the algorithm. This measure was designated as the Self-Consistency Error or S.C.E. It is defined mathematically in the Appendix.

Two other measures were developed to assess the accuracy of determination of projection directions more directly. Although Euler angles were used to define the projection directions we could not use them to quantify accuracy of alignment directly, since widely different Euler angles may specify similar projection directions. Instead, the final positions of the x and z axes of the aligned projection images were compared with the x and z axes corresponding to the initial projection directions used to generate the test images. The average angular discrepancy between these pairs of axes was defined as the Coordinate System Error or C.S.E. (see Appendix, equation A.3).

The C.S.E. does not, however, discriminate between individual misalignments and a general rotation of all the projection systems. Since accurate 3D reconstruction requires only that relative projection directions be determined correctly, we developed a third measure. This measure is referred to as the Residual System Error, R.S.E. (see Appendix equation A.4). The R.S.E. is essentially similar to the C.S.E. but an optimal general rotation (Appendix equation A.5) is applied to the projection coordinate systems described by the experimentally determined angles.

4.2. Experimental Simulations

The Klenow fragment of DNA polymerase I was chosen to study the above techniques. The Klenow fragment has properties that make it a suitable candidate to determine how effectively the IQAD method may be applied to E.M. data. The 68kD molecular weight of the Klenow fragment provides sufficient contrast for easy detection in dark-field electron microscopy of the unstained molecule. Lack of symmetry in the structure results in a specimen not easily
amenable to other techniques. The availability of the X-ray crystallographic structure allows the simulations described below to be performed and provides the standard for comparison with an electron microscopy-derived structure.

A volume representation of the Klenow fragment of DNA polymerase I was generated at a 0.54 nm resolution from the published positions of the alpha carbons of the protein backbone. To create a more realistic representation of the structure, the mass corresponding to each amino acid was distributed in a spherical approximation over the 26 voxels surrounding the voxel that contained the position of the corresponding alpha carbon. Because the side of each voxel, 0.27 nm, is greater than half the average distance between vicinal alpha carbon atoms, the dispersion of mass over neighbouring voxels results in a continuous mass distribution in the volume.

Simulated E.M. images of the Klenow fragment were produced by forward projection of the volume in the manner described by Harauz. All projection images generated were interpolated onto a 64 x 64 pixel array. A single data set contained 20 groups of images each at different projection directions defined by randomly chosen Euler angles, denoted (α,β,γ). Each group had 5 members providing a data set with a total of 100 projections. Projection images within a group were either identical or were perturbed by noise, in angular orientation or both.

Six different sets of simulated data were presented to the IQAD algorithm. In the first simulation, Sim(i,∞), the projection directions within a group were identical, (symbol “i”) and no noise was added to the images (signal to noise ratio = ∞). In the second and third simulations, the projection directions within a group were again identical but Gaussian distributed noise with zero mean was added to the images to give a signal to noise ratio of 6.2 and 3.1 respectively. These two simulations were denoted Sim(i,6.2) and Sim(i,3.1) using a similar convention. The signal to noise ratio of 6.2 was chosen, as it corresponds to the experimentally determined signal to noise ratio for a typical set of our dark-field electron microscope images of the Klenow fragment. In the final three simulations the images within a group were projected with a random angular perturbation (symbol “p”) up to ±10° in α and γ and ±5° in β. The first of the three sets of angularly perturbed images, Sim(p,∞), had no additional Gaussian noise. The second, Sim(p,6.2) and third, Sim(p,3.1) had signal to noise
The mean R.F.A. error is the root mean square deviation within the original groups of images in degrees.

Table 1. Projection Data set description and reference-free alignment (R.F.A.) results.

<table>
<thead>
<tr>
<th>Sim (i,∞)</th>
<th>Sim (i,6.2)</th>
<th>Sim (i,3.1)</th>
<th>Sim (p,∞)</th>
<th>Sim (p,6.2)</th>
<th>Sim (p,3.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal to Noise Ratio</td>
<td>∞</td>
<td>6.2</td>
<td>3.1</td>
<td>∞</td>
<td>6.2</td>
</tr>
<tr>
<td>Perturbed Projections</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Mean R.F.A. σ</td>
<td>0.00</td>
<td>0.50</td>
<td>1.03</td>
<td>10.65</td>
<td>10.70</td>
</tr>
</tbody>
</table>

A reference-free alignment was performed, each image being aligned with respect to the composite image over 5 iterations. The aligned images were then classified using the multivariate statistical analysis packages available in the IMAGIC software routines. The following parameters that define the classification protocol were the result of a number of trials with similar sorts of experimental conditions as those used in this experiment. A Euclidian measure of distance was used to determine relative positions of images in pixel space. Twenty iterations were used to determine 10 eigenimages that best described the 100 images in the experiment. A modified hierarchical ascendent classification scheme was used to separate the 100 images into 20 classes. Members of each class were then summed to increase the signal to noise ratio of the images passed to the 3D orientation determination process.

Sinograms, a collection of line projections at 1° intervals, were generated from all the averaged projection images. The IQAD procedure determined which sinograms would be compared using sinogram correlation functions to permit the determination of common line projections between different images. From the positions of the common lines the relative orientations of the projections were determined as described in Chapter 3.

Examination of problems with the alignment of some of the data sets led to the modification of the IQAD alignment process. This modified process, referred to as mIQAD.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
is described below in the context of the results from Sim(i,3.1). The modified procedure was used to re-determine the projection directions for all the simulated data sets.

All computations were performed on a pVAX II (Digital Equipment Corporation, Maynard, MA, USA). Application of the reference-free alignment technique to 100 projection images required approximately 9 hours for alignment. The determination of eigenimages and classification required 20 minutes. Generation of sinograms, calculation of sinogram correlation functions and determination of projection directions using IQAD techniques required approximately 10 hours for 20 images. Calculation of individual sinograms requires 10 minutes and the calculation of a sinogram correlation function requires 4 minutes on this computer.
4.3 Results and Discussion

4.3.1 Definitions

In the following discussion, images that were originally generated with a common projection direction are referred to as members of an image "group". Images that are similar as defined by principal component analysis will be referred to as members of a "class". Image groups are referred to by a letter, A to T. Individual group members are referred to by a number, hence group A contains images A1 to A5. Classes which do not correspond exactly to a group are denoted by an asterisk "*".

4.3.2 Reference-Free Alignment Results

The reference-free alignment technique was implemented to bring similar images into registration with one another. The success of this procedure can be gauged by examining the variation in the rotation angle used to align images within a group to the data set as a whole. The standard deviation of the rotation angle was calculated for each group in each data set and the average intra-group standard deviation for each data set is given in Table 1.

In the data set Sim(i,∞) all images within each of the twenty groups are rotated through exactly the same angle. In Sim(i,3.1) the rotation angles have an average intra-group standard deviation of 1.0°. The angular deviations for the simulations which contain images at perturbed projection directions in each group are somewhat higher. This is not unexpected, as the initial γ perturbation alone would be expected to give a standard deviation of 5.7°. In one group of images in both Sim(p,∞) and Sim(p,6.2) a single image out of the 100-member set was rotated approximately 180° with respect to the other members of that group. When these single images were removed from the analysis the average intra-group standard deviation of the rotation angle was 6.8° for both Sim(p,∞) and Sim(p,6.2). In Sim(p,3.1) two such obvious misregistrations occurred. If the misrotated images are removed from this analysis the average intra-group standard deviation of the rotation angle falls to 7.0°.
Figure 2. Examples of the misregistration of images that occurred as a result of the reference-free alignment procedure. The images of group M from Sim(p, oo) are shown in the upper row. Image M2 is rotated by 180° with respect to the other images in the group. The lower row shows the images from group J of Sim(p,3.1). Images J1 and J3 are rotated by a 120° counter-clockwise rotation with respect to the other members of the group.

These misrotations, that occurred in groups M and J could not be connected to any major, visually detectable differences in the projections, (see Fig. 2). However, Group M contained one of the most compact and indistinct views of the molecule.

4.3.3 Classification Results

Once the images had been aligned using the reference-free alignment algorithm principal component analysis based techniques were used to assign similar images to individual classes. When the images from Sim(i, oo) and Sim(i,6.2) were classified no errors occurred, i.e. the members of a class corresponded exactly to the members of one of the original groups. In all of the other simulations some classes differed from the original groups of projections. These differences are listed in Table 2.
Table 2. Classification results. Only classes whose members do not correspond to an original group of images in the simulation are shown.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Class Designation</th>
<th>Class Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Sim}(i,3,1) )</td>
<td>( A' )</td>
<td>( A1, A2, A4, A5, G3 )</td>
</tr>
<tr>
<td></td>
<td>( F' )</td>
<td>( F1, F2, F3, F5, R5 )</td>
</tr>
<tr>
<td></td>
<td>( G' )</td>
<td>( A3, G1, G2, G4, G5 )</td>
</tr>
<tr>
<td></td>
<td>( R' )</td>
<td>( F4, R1, R2, R3, R4 )</td>
</tr>
<tr>
<td>( \text{Sim}(p, \infty) ) and ( \text{Sim}(p,6.2) )</td>
<td>( A' )</td>
<td>( A1, A2, A3, A5, G1, G2, G3, G4, G5 )</td>
</tr>
<tr>
<td></td>
<td>( F' )</td>
<td>( F1, F2, F3, F4, F5, R1, R2, R3, R4, R5 )</td>
</tr>
<tr>
<td></td>
<td>( G' )</td>
<td>( K3, K5 )</td>
</tr>
<tr>
<td></td>
<td>( K' )</td>
<td>( K1, K2, K4 )</td>
</tr>
<tr>
<td></td>
<td>( O' )</td>
<td>( O1, O2, O5 )</td>
</tr>
<tr>
<td></td>
<td>( R' )</td>
<td>( O3, O4 )</td>
</tr>
<tr>
<td>( \text{Sim}(p,3.1) )</td>
<td>( A' )</td>
<td>( A1, A2, A3, A4, A5, G1, G2, G3, G4, G5 ) ( N1, N2, N3, N4, N5 )</td>
</tr>
<tr>
<td></td>
<td>( D' )</td>
<td>( D1, D2, D3, D4, D5, J5 )</td>
</tr>
<tr>
<td></td>
<td>( F' )</td>
<td>( F1, F2, F3, F4, F5, R1, R2, R3, R4, R5 )</td>
</tr>
<tr>
<td></td>
<td>( G' )</td>
<td>( J1, J3 )</td>
</tr>
<tr>
<td></td>
<td>( I' )</td>
<td>( I1, I2, I3, J4 )</td>
</tr>
<tr>
<td></td>
<td>( J' )</td>
<td>( J2, J4, J5 )</td>
</tr>
<tr>
<td></td>
<td>( K' )</td>
<td>( K1, K2, K4 )</td>
</tr>
<tr>
<td></td>
<td>( L' )</td>
<td>( L1, L2, L3, L4 )</td>
</tr>
<tr>
<td></td>
<td>( N' )</td>
<td>( O3, O4 )</td>
</tr>
<tr>
<td></td>
<td>( O' )</td>
<td>( O1, O2, O5 )</td>
</tr>
<tr>
<td></td>
<td>( Q' )</td>
<td>( L5, Q1, Q2, Q3, Q4, Q5 )</td>
</tr>
<tr>
<td></td>
<td>( R' )</td>
<td>( K3, K5 )</td>
</tr>
</tbody>
</table>
The misclassifications in Sim(i,3.1) result from similarities in the Euler angles used to define different groups of projections. By chance the Euler angles used to define projection groups F and R are (79.8, 34.4, 275.7) and (72.1, 43.1, 282.5) respectively (allowing for the rotation introduced by the reference-free alignment process). As the Euler angles are very similar, one would expect the corresponding projection images to be similar. The Euler angles defining projection groups A and G are (0.0, 0.0, -0.4) and (231.1, 7.3, 130.0) respectively (after allowing for the reference-free alignment process rotation). Despite the large differences in the value of the Euler angles the resulting rotations are almost identical. Examples of the images in these classes are shown in Fig. 3.

The classification procedure gave identical classes for both Sim(p,∞) and Sim(p,6.2). In these simulations class A* is formed by merging groups A and G and class F* is formed by merging groups F and R: such a merging of classes occurs because, as discussed above, the projection directions of the groups are almost equivalent.

The classification procedure partitions group K into classes K* and G* and group O into classes O* and R*. Examination of the images in group O (Fig 3.) shows that the angular perturbation of the views has resulted in two visually different types of projections. The two views are partitioned in agreement with the distribution in classes O* and R*. A similar although less striking observation may be made for the images in group K (Fig. 3). Interestingly the images in group M are assigned to a single class despite the misrotation resulting from the reference-free alignment procedure.

As expected, application of classification procedures to the data in Sim(p,3.1) resulted in greater misclassification. In general, the reasons are similar to those described above for groups F and R, groups A, G and N, groups I and D, and groups K and O (see Fig. 3 and Table 2). Group J is split into two classes as a result of the different rotations introduced within the group by the reference-free alignment procedure (see Fig. 2). Finally image L5 is merged with group Q. The Euler angles of groups L and Q, (224.6, 85.9, 195.0) and (210.8, 43.1, 282.4), do not appear particularly close, but inspection of the images reveals the projections to be similar (see Fig. 3).
Figure 3. Examples of images which were not classified according to their original groups, all images are shown after reference-free alignment rotation. Images F and R (from Sim(p,3.1)) are examples of their groups which were merged to form class F. Similarly images A and G were merged to form class A. Image N is an example of the group N images which were merged with groups A and G in Sim(p,3.1). The members of image groups K and O from Sim(p,10) illustrate the groups that were partitioned into separate classes. Group K formed two classes with members (K1, K2, K4) and (K3, K5) whilst group O was split into two classes with members (O1, O2, O5) and (O3, O4). Images L5 and Q, an example of group Q, from Sim(p,3.1) indicate the similarity between the two projections despite diverse Euler angles.
4.3.4 Projection Direction Determination Results

The classified images were processed by the IQAD algorithm to determine the relative orientations of the projections. The resulting orientations were examined using the three measures described in the Appendix (section 4.A.2). The results are given in Table 3.

The IQAD algorithm was able to align the projections from Sim(i,∞) and Sim(i,6.2) to an accuracy approximately equivalent to the 1° interval with which the positions of the common axes were determined. All three measures of algorithm performance indicate this to be the case. The projection directions determined using the images from Sim(i,3.1) are incorrect however. The experimentally determined projection directions are compared to the correct directions in Fig 4.

The figure represents the projection directions by indicating the positions at which the projection direction vectors would intersect with the surface of a sphere. These intersections must be represented in two dimensions, this problem is akin to that faced by cartographers when describing the Earth on a flat piece of paper. We have developed our own representational system and this system is used to represent projection directions throughout the rest of this thesis. A projection direction may be described by the two Euler angles α and β which define a position on the surface of a sphere in the same way that a latitude (β) and longitude (α) define a position on the Earth. However it is not necessary to describe a full

Table 3. The self consistency error, coordinate system error and residual system error of alignments using the IQAD procedure. All errors are expressed in degrees, the parentheses contain the standard deviation of the value.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>S.C.E.</th>
<th>C.S.E.</th>
<th>R.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sim(i,∞)</td>
<td>0.9(0.1)</td>
<td>1.1(1.2)</td>
<td>0.9(1.2)</td>
</tr>
<tr>
<td>Sim(i,6.2)</td>
<td>1.7(1.1)</td>
<td>2.3(2.5)</td>
<td>2.0(2.2)</td>
</tr>
<tr>
<td>Sim(i,3.1)</td>
<td>19.6(25.0)</td>
<td>19.0(29.3)</td>
<td>19.8(26.3)</td>
</tr>
<tr>
<td>Sim(p,∞)</td>
<td>0.7(0.7)</td>
<td>4.3(3.7)</td>
<td>3.3(3.2)</td>
</tr>
<tr>
<td>Sim(p,6.2)</td>
<td>3.8(16.0)</td>
<td>6.1(5.8)</td>
<td>4.9(5.0)</td>
</tr>
<tr>
<td>Sim(p,3.1)</td>
<td>20.2(21.9)</td>
<td>18.2(22.6)</td>
<td>18.3(20.1)</td>
</tr>
</tbody>
</table>
Figure 4. Projection directions of the classes of images from the data set Sim(i,3.1). The projection directions are specified by the Euler angles $\alpha$ and $\beta$. These angles denote the position at which the projection direction vector intersects the surface of a hemisphere (see description in section 4.3.4). The plot represents the surface of the hemisphere in two dimensions. The figure is similar to a conventional polar plot, $\beta$ is plotted in the radial direction whilst $\alpha$ is indicated by the angular displacement around the circle. The projection direction of each class is indicated by its corresponding letter. The lines extending from each symbol indicate the displacement of each class from its expected position. The "***" indicates classes of images not corresponding exactly to an original group of images, see section 4.3.1.

sphere because projections at angles $\alpha, \beta$ are essentially equivalent to projections at $\alpha+180^\circ, 180^\circ-\beta$. Views at these two directions would contain the same information collected from opposite directions. For this reason the figure represents only the surface of a hemisphere with $\alpha$ running from $0^\circ$ to $360^\circ$ and $\beta$ running from $0^\circ$ to $90^\circ$. The plot is essentially similar to a conventional polar plot in which $\beta$ is plotted in the radial direction whilst $\alpha$ is indicated by the angular displacement around the circle. Projection directions with
a value of $\beta$ less than $90^\circ$ are indicated by an open circle. Projection directions with a value of $\beta$ between $90^\circ$ and $180^\circ$ are displayed at their equivalent position on the hemisphere, i.e. $\alpha+180^\circ$, $180^\circ-\beta$ and this transformation is indicated by a solid symbol.

The petal-shaped cross pattern in the centre of the figure arises for the following reasons. The length of a line of latitude (constant $\beta$) on the surface of a sphere is proportional to $\sin(\beta)$. The boundaries of the quadrants of the hemisphere are modified to reflect this relationship restricting the area in which projection directions may be plotted. The boundaries of the modified quadrants represent values of $\alpha$ of $0^\circ$, $90^\circ$, $180^\circ$ or $270^\circ$.

The errors in the determination of the projection directions are indicated by lines which run from the experimentally determined projection direction positions to the corresponding expected projection direction position for each class of images.

From Fig. 4 it is apparent that there are some large errors in the determination of the projection directions for the data set Sim(i,3.1). The projection directions determined for classes O, P and Q are seen to be those with the greatest error.

From Table 3 it is seen that the experimentally determined projection directions from Sim(p,$\infty$) and Sim(p,6.2) have errors that are greater than those in the equivalent noise level projections in Sim(i,$\infty$) and Sim(i,6.2). However, the assigned projection directions for these data sets deviate by only a few degrees from their correct values. Only in the low signal to noise ratio data of Sim(p,3,1) does the quality of the alignment fall significantly.

Problems aligning high noise images indicated modification of the IQAD algorithm was required. To understand the principle behind the modification it is necessary to examine the individual angular discrepancies between pairs of corresponding axes when they are aligned in three dimensions. The "spectrum" of these angular errors, referred to as individual S.C.E.s, is plotted in Fig. 5. The spectrum illustrates the individual contribution of each common axis to the S.C.E. The figure shows that for the majority of common axes the alignment procedure has succeeded in determining a nearly coincident position of the common axis pairs in the 3D coordinate system. However for certain pairs there is a significant difference between the two calculated positions of the individual common axis, e.g. in the data set in Fig. 5, 24% of the positional discrepancies are greater than $25^\circ$.

The outlying points could arise either from poor alignments of the projections or as
a result of poor input data, i.e. the positions of the common axes between a pair of projections were wrongly determined. Examination of the positions of the common axes revealed the latter explanation to be correct.

Information similar to that illustrated in Fig. 5 may be determined during the alignment procedure. Therefore a modification was made to the algorithm to allow it to reject the data contributing to the outlying points. The modified algorithm proceeded as follows. An alignment was performed that would yield a spectrum of individual S.C.E.s similar to that illustrated in Fig. 5. The individual S.C.E.s were examined to determine which common axis corresponded to the largest individual S.C.E. The data associated with this value were then removed and a fresh alignment performed without the putatively bad data point. If the overall S.C.E. of the new alignment improved as a result of leaving out the data point then the process was repeated removing the common axis data corresponding to the new, worst individual S.C.E. When removal of a data point did not improve the overall value of S.C.E. the previous alignment was accepted as defining the final orientations. This modified IQAD
The self consistency error, coordinate system error and residual system error of alignments using the mIQAD procedure. All errors are expressed in degrees, the parentheses contain the standard deviation of the value. The number of common axes used for the final alignment (out of an original 107) is given as N.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>S.C.E.</th>
<th>C.S.E.</th>
<th>R.S.E.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sim(i,∞)</td>
<td>0.5(0.2)</td>
<td>1.2(0.7)</td>
<td>0.8(0.4)</td>
<td>81</td>
</tr>
<tr>
<td>Sim(i,6.2)</td>
<td>1.0(0.4)</td>
<td>2.0(0.9)</td>
<td>1.2(0.9)</td>
<td>87</td>
</tr>
<tr>
<td>Sim(i,3.1)</td>
<td>1.3(0.6)</td>
<td>12.8(35.4)</td>
<td>10.7(34.6)</td>
<td>67</td>
</tr>
<tr>
<td>Sim(p,∞)</td>
<td>0.2(0.1)</td>
<td>4.7(2.1)</td>
<td>2.9(2.2)</td>
<td>71</td>
</tr>
<tr>
<td>Sim(p,6.2)</td>
<td>0.7(0.3)</td>
<td>5.6(2.2)</td>
<td>3.1(2.3)</td>
<td>81</td>
</tr>
<tr>
<td>Sim(p,3.1)</td>
<td>1.3(0.6)</td>
<td>9.7(16.9)</td>
<td>7.0(17.8)</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 4. The self consistency error, coordinate system error and residual system error of alignments using the mIQAD procedure. All errors are expressed in degrees, the parentheses contain the standard deviation of the value. The number of common axes used for the final alignment (out of an original 107) is given as N.

The use of the mIQAD procedure improves the S.C.E. for all simulations. Examination of the alignment errors for Sim(i,∞) and Sim(i,6.2) reveals an extremely close agreement between the calculated projection directions and those used to generate the simulation data (R.S.E.s of 0.8° and 1.2° respectively). The calculated projection directions for Sim(i,3.1) are plotted in Fig 6. It can be seen that all the projection directions, with the exception of class R*, have been determined correctly. This error is discussed below.

The projection direction assigned to projection R* of Sim(i,3.1) using the mIQAD procedure is (267.89, 149.92, 253.12). The correct direction used in the simulation was (72.07, 43.13, 283.31). A detailed examination of the alignment history of this projection reveals that the original data from the sinogram correlation functions had significant errors. In all, 9 common lines were calculated between class R* and other projections. Three of the common axis positions were in error by greater than 120°. Since the algorithm initially treats all data points as valid, the calculation of the wrong common lines via the sinogram correlation function leads to the incorrect determination of the projection directions. When this incorrect determination was made the erroneous common axes were, by happenstance,
Figure 6. miQAD determined projection directions for the data in Sim(i,3.1). Only the data for class R is in error.

in very good agreement, (as defined by small individual S.C.E.s). Those for which the common axis data were determined more correctly had greater individual S.C.E.s. Because the three misdetermined common axes can be so successfully aligned the other "good" data points were removed by the miQAD algorithm.

Considering Sim(p,∞), Sim(p,6.2) and Sim(p,3.1), the quality of the alignment decreases (as expressed by all three measures) as the noise in the data increases. Nevertheless, even at a signal to noise ratio of 3.1 the perturbed data is aligned with considerable accuracy, see Fig. 7.

Comparison of Tables 3 and 4 reveals that the use of the miQAD algorithm, according to design, reduces the S.C.E. from its value given by IQAD. A similar improvement is not seen in general for the measures C.S.E. and R.S.E. However, in none of the simulations is a serious degradation of the C.S.E. and R.S.E. found as a result of implementation of the
Another interesting observation can be made with respect to the IQAD solution of the Sim(i,3.1) data. Contrary to what one would expect, the value of the R.S.E. is slightly higher than that of the C.S.E. This is a result of expressing the similarity between the systems as an average angle, whilst the optimum global rotation is calculated in terms of the average scalar product (i.e. the cosine of the angles) between the x and z axes of the coordinate systems.

In extrapolating the above simulation results to experimental situations it should be remembered that the simple noise model considered above may only represent part of the degradation of image quality that occurs in real electron microscope images. Damage to the specimen from the electron beam,\textsuperscript{37,38} mobile portions of the molecule and differential forces from specimen preparation\textsuperscript{39} will all serve to further degrade image quality and complicate the above procedure.
4.4 Conclusions

This study has addressed the problems inherent in determining the projection directions of a series of images \textit{a posteriori}. The problem was broken down into three separate stages, reference-free alignment, image classification and alignment determination.

The results described above are specifically related to the structure of the molecule chosen for this series of simulations, it is expected that results might vary somewhat for different molecular structures depending on properties of the structure, e.g. its degree of asymmetry. However, the results described above have usefully demonstrated potential problems with the method as well as their solution and have led to predictions of the accuracy of the alignments that might be achieved using these methods.

It has been shown that reference-free alignment facilitates similar rotation of similar images in simulations in which images within a group differed by only the noise that they contained. When the images were perturbed, differences in rotation were attributed to different orientations giving rise to slightly different projection images. However, even at the highest noise level and with angular perturbation of the projection images only 4% of images were rotated to positions significantly different from other members of their respective group.

Principal component-based procedures were used to classify similar images. No classification errors occurred for non-perturbed images when no noise was added to the image, or at a signal to noise ratio of 6.2. At a signal to noise ratio of 3.1 only minor classification errors were made. The errors that arose were due to the presence of similar images in different groups. When the angle-perturbed image simulations were considered, the additional misclassifications could all be ascribed either to differential rotation of images under the reference-free alignment process or to the presence of different views within a group.

The use of the mIQAD procedure allowed the effective determination of projection directions even in the presence of combinations of noise and angle perturbed images within a single group. Only one significant error was made, group $R^*$ in Sim(i,3.1). To detect such erroneous alignments the mIQAD procedure should be implemented a number of times and the consistency of the relative projection directions in the different experiments should be

126
examined. In Sim(p,3.1) the average misalignment of the coordinate systems of the individual projections was only $7^\circ$.

In summary it has been shown that despite the cascade of procedures necessary to increase the quality of images prior to the determination of orientations it is possible to determine the relative projection directions of a large number of projection images \textit{a posteriori} to an accuracy that augurs well for their use in 3D reconstruction.

Acknowledgements

We thank Marin van Heel and George Harauz for the use of the IMAGIC image processing package and George Harauz for his many helpful comments. This work was funded by the Medical Research Council of Canada, the National Cancer Institute of Canada and the Ontario Cancer Research and Treatment Foundation.
Appendices

In section 4.1.2 the method by which the relative orientation of the projection images may be determined was described. In section 4.A.1 below the same principles are expressed using a mathematical formalism. Similar mathematical descriptions, (section 4.A.2) may then be employed to define the three measures of algorithm performance described in section 4.1.3.

4.A.1 Definition of Optimum Alignment

Consider the alignment of projection $j$. Let projection $j$ have common axes with $N$ other projections denoted $i$, \(i = 1, N\). The orientations of these $N$ projections with respect to a common 3D coordinate system will have been calculated previously and be defined by their 3D rotation matrices $T_i$. The position of the common axis between a projection $i$ and projection $j$ in the coordinate system of projection $i$ is denoted by the vector $c'_{ij}$. The position of this common axis in the common 3D coordinate system is given by $c'_{ij}T_i$ and is denoted $c_{ij}$. Vectors $c'_{ij}$ denote the position of the common axes between projections $i$ and projection $j$ in the coordinate system of projection $j$. Determination of the optimum orientation of projection $j$ relies on the fact that the vectors $c'_{ij}$ under a rotation denoted by matrix $T_j$ should coincide with vectors $c_{ij}$ in the common coordinate system, i.e. $c_{ij} = c'_{ij}$. The rotation matrix that describes the orientation of projection $j$ is given by determining the matrix $T_j$ that maximizes the following expression

$$\sum_{i=1}^{N} c'_{ij}T_i \cdot c'_{ij}T_j$$

The matrix $T_j$ which maximizes the above expression may be determined by Quaternion mathematics.\(^{33,34}\)


To evaluate the performance of the algorithm a number of measures have been developed.
The first measure, the Self-Consistency Error or $S.C.E.$, represents the average angular difference between the positions of the vectors $c_i$ and $c_i'$ in the common 3D coordinate system. The $S.C.E.$ is written explicitly as

$$S.C.E. = \frac{1}{N_c} \sum_i \sum_j \arccos( c'_i T_i \cdot c'_j T_j )$$

where $N_c$ is the total number of common axes considered in the data set. The $S.C.E.$ quantifies how successfully the algorithm has been able to use the input data (in the form of common axes positions in the projection coordinate systems) to derive a self-consistent 3D alignment. For all measures of algorithm performance "unsigned" inverse cosine functions are used to effectively yield the magnitude of the angular discrepancies.

Caution must be exercised in the interpretation of the $S.C.E.$ parameter as an indicator of the success of the alignment per se. If the original angular positions of the common axes contain errors, as will be likely for data from noisy projections, then the $S.C.E.$ will to some extent describe a combination of both the errors in the input data and the success of the alignment. In other words a distinction cannot be made between the correct orientation of projections in which the initial data was in error and incorrect orientation of projections in which the input data was correct. In both cases vectors $c_i$ and $c_i'$ will not be equivalent and the discrepancy between positions of the vectors will be recorded in the $S.C.E.$

To assess the accuracy of determination of projection directions more directly, two other measures were developed. Although Euler angles were used to define the projection directions we do not use them to quantify accuracy of alignment directly, as widely different Euler angles may specify similar projection directions. Instead we compare the positions of the $x$ and $z$ axes (expressed as vectors $x$ and $z$) of the projection coordinate systems rotated by their experimentally determined $T$ matrices to the positions the same axes vectors would adopt as a result of rotation by the matrices used to generate the initial projection images, denoted $U$. A second measure, the coordinate system error, $C.S.E.$, was then used to define the average angular discrepancy between these pairs of axes. Explicitly,
\[ C.S.E. = \frac{1}{2N_p} \sum_{i=1}^{N_p} \left( \arccos(xT_i \cdot xU_i) + \arccos(zT_i \cdot zU_i) \right) \]  

A.3

where \( N_p \) is the total number of projections under consideration.

The \( C.S.E. \) does not however discriminate between individual misalignments and a general rotation of all the projection systems. Since, accurate 3D reconstruction requires only that relative projection directions be determined correctly, we developed a third measure. This measure is referred to as the residual system error, \( R.S.E. \), and is defined

\[ R.S.E. = \frac{1}{2N_p} \sum_{i=1}^{N_p} \left( \arccos(xT_iT_g \cdot xU_i) + \arccos(zT_iT_g \cdot zU_i) \right) \]  

A.4

The measure is essentially similar to that described by the \( C.S.E. \) but an optimal general rotation is applied to the projection coordinate systems described by the experimentally determined angles. The matrix describing this general rotation, \( T_g \), is determined by an extension of the Quaternion mathematics described by Horn.\(^3\) The matrix maximizes the expression

\[ \sum_{i=1}^{N_p} \left( (xT_iT_g \cdot xU_i) + (zT_iT_g \cdot zU_i) \right) \]  

A.5

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


CHAPTER 5

THREE-DIMENSIONAL STRUCTURAL STUDIES OF THE KLENOW FRAGMENT OF DNA POLYMERASE I USING DARK-FIELD ELECTRON MICROSCOPY

Neil A. Farrow and F. Peter Ottensmeyer

In previous chapters a method has been described which permits the \textit{a posteriori} determination of the relative orientations of 2D projection images of a 3D density distribution. Knowledge of such orientations permits the reconstruction of 3D density distributions from the 2D images. In this chapter dark-field electron micrographs of the Klenow Fragment of DNA Polymerase I are used to investigate the 3D structure of the molecule. The orientation of the molecule in each image is initially unknown and is determined using the common axis theorem based techniques described previously. The image processing steps that are required for such a study are discussed in detail. A 2D reference-free alignment procedure is implemented successfully and images are classified using principal component analysis techniques to permit image enhancement by averaging similar images. Examples of the reconstructed volumes derived from the electron micrographs are shown. The 3D reconstructions are rather featureless and do not resemble the structure previously determined by X-ray crystallography. Recognizable problems with both the initial images and the alignment technique are discussed. In the context of those conditions, a possible explanation for the low quality 3D reconstruction is given.
5.1 Introduction

In Chapter 3 an algorithm was described that permitted the determination of the relative 3D orientations of a series of projection images of objects at unknown random orientations. The algorithm employs the principles of the common axis theorem\(^1\) and quaternion mathematics\(^2^{,3}\) to align the images with respect to one another. The data that the algorithm requires to perform such alignments consists of the angular positions of common line projections between pairs of images. In Chapter 3, these positions were calculated analytically in a manner that required only the specification of a series of test projection directions. In Chapter 4 the positions of the common line projections were determined from simulated projection images using sinogram correlation functions.\(^4\) The simulations described in Chapter 4 demonstrated that the algorithm was capable of determining the projection directions of simulated noisy, perturbed images of the Klenow Fragment of DNA Polymerase I.

The preceding simulations were designed to model the data that would be presented to the alignment algorithm when the latter were applied to real dark-field electron micrographs. Following the successful application of the algorithm to simulated data it was decided to investigate the efficacy of the algorithm when applied to real dark-field electron micrographs of the Klenow Fragment. The results of this investigation are described in this chapter along with the practical aspects of applying such a technique to a collection of digitized images.

In the same way that the simulations described previously were able to establish that the alignment algorithm was working correctly, the experiments described here, using images of the Klenow Fragment, were designed to demonstrate that the alignment algorithm would perform correctly when applied to real electron micrographs. It was for this reason that the Klenow Fragment was chosen as a molecule suitable for this investigation. The structure of the Klenow Fragment has previously been determined using X-ray crystallographic techniques.\(^5\) It was then anticipated that the structure obtained from the use of the alignment algorithm and dark-field electron micrographs would, to some extent, resemble that structure previously determined. If this were the case then future application of the algorithm, to determine other unknown molecular structures, could be justified.
In Chapter 1 a number of factors which contribute to the degradation of electron microscope images were discussed. The primary source of concern related to damage to the specimen which results from its interaction with the electron beam. The use of a scanning transmission electron microscope (STEM) is one possible technique that may be applied to reduce the electron dose received by the specimen. This reduction occurs as a consequence of the five-fold increase in electron collection efficiency of the STEM over a conventional transmission electron microscope. Prior to the studies described below, we investigated the possibilities of using the STEM at McMaster University in Hamilton. Use of this instrument was rejected for a number of reasons. The main problem concerned the poor quality of the Polaroid negatives that were used to record the images; the facility in Hamilton was primarily concerned with X-ray microanalysis studies in materials science and the STEM had not been optimized for the type of imaging studies that we were proposing.

An additional approach to reducing the effects of radiation damage is to image the specimen at low temperatures. It has been estimated that the use of a specimen stage at liquid helium temperature might increase the critical dose for some damage end point by as much as a factor of 5.0.6 Liquid nitrogen temperatures are not quite as efficacious in their protective effect, but are still very useful.6 A low temperature specimen stage operating at liquid nitrogen temperatures was available and was used very recently by members of our group at the Brookhaven National Laboratories, Long Island, New York, to image the Klenow Fragment in the STEM. Due to low electron currents from the field emission source of that microscope, images had signal to noise ratios that were too low for useful processing. STEM imaging could be carried out with sufficient signal to noise ratio at the Müller Institute in Basel, Switzerland though without a low temperature stage. Unfortunately, during a brief work period which was arranged as part of a European visit, specimen preparation problems prevented useful images from being recorded. Nevertheless, dark-field images using a STEM at low temperatures will continue to be pursued, since we consider these as the best images for the processing and reconstruction algorithm.

In the meantime we have obtained images of the Klenow Fragment using tilted beam dark-field in a Philips EM 300 electron microscope. It is recognized that these images do not represent the highest quality dark-field electron micrographs that could theoretically be
obtained. Nevertheless, previous studies have demonstrated that structures on the order of 0.5 nm were resolvable on smaller molecules using techniques similar to those employed in the experiments described in this chapter.\textsuperscript{7,8} Moreover, a number of dark-field images of the Klenow Fragment clearly showed encouraging similarities with recognizable orientations of the X-ray structure of the molecule (see Fig. 2, Chapter 1). For these reasons, it was decided to test the practical applicability of the alignment algorithm using this type of dark-field micrograph. The proposed study would be beneficial to future practical applications of the algorithm in that it might elucidate unforeseen factors that were not included as part of the previously described simulations.

A number of image processing techniques are available to increase the quality of the input images to the alignment procedure. These techniques were discussed in Chapter 1. The technique that is employed in the following investigation is that of averaging similar images.\textsuperscript{7,9-12} Averaging of similar images first requires the identification and classification of similar images. Because of the large amount of information contained in a single image, the dimensionality of the data set must first be reduced. This reduction is achieved by the application of principal component analysis techniques.\textsuperscript{13,14} Such techniques have been widely applied in bright-field electron microscopy studies,\textsuperscript{15-17} but to our knowledge have not been applied previously to dark-field electron micrographs. Following this data reduction step, the images may then be automatically classified prior to enhancement via averaging.\textsuperscript{18,19}

The application of these analysis and classification techniques to dark-field microscope images is described in this chapter. Also in this chapter, the results of applying the reference-free alignment technique (described in Chapter 4) to dark-field images are described. Section 5.3 describes the 3D reconstructions of the Klenow Fragment that result from applying the alignment algorithm to dark-field electron micrographs. It is seen that the 3D reconstructions do not bear a great resemblance to the structure that was determined previously by X-ray crystallography. At best, an indication of a grooved structure is seen that hints at the location of the very prominent DNA binding cleft of the X-ray structure. In section 5.4 the discrepancy between the two structures is discussed in the context of the results that are obtained at intermediate stages from both the reference-free alignment procedure and the sinogram correlation functions.
5.2 Methods

5.2.1 Electron Microscopy and Digitization

To prepare the sample for electron microscopy 5 µl of DNA polymerase I (Pharmacia FPLCpure™) at an original concentration of 640 µg/ml was added to 59 µl of buffer (130 mM potassium phospbate, 6.5 mM magnesium chloride, 1 mM dithioerythreotol, pH 7) to give a final concentration of 50 µg/ml. A 10 µl drop of sample was then applied to an ultrathin, indirectly evaporated carbon film supported on a fenestrated film of cellulose acetate butyrate on a 200 mesh copper grid. The sample was allowed to adsorb for 3 minutes and was then rinsed to remove excess sample and buffer using 5 drops of deionized, distilled water. Excess liquid was removed from the surface of the grid by touching a damp filter paper to the grid. The sample was then allowed to air dry.

Electron micrographs of the molecule were obtained using the minimum beam exposure technique in which a neighbouring area of the sample is used to focus the microscope, the beam is then deflected electronically, the sample is moved to a new area and the beam deflection removed as the micrograph is taken. All images were taken using a Philips EM 300 microscope operating at 80 kV in tilt beam dark-field mode. The calibrated magnification used was 36,900 x. Images were collected on glass plates (Electron Image Plates, Kodak) and were developed using standard conditions. The optical density of the carbon support film in the images plates was = 0.35 O.D., ensuring the linear range of the plate was used. From this O.D. value it was estimated that the exposure of the sample (assuming a 3 nm thick support film) was approximately 1000 eÅ².

Images were examined using a low power dissecting microscope to check focus, astigmatism and lack of drift. From the 60 micrographs that were acquired, 5 were selected for digitization. Digitization was performed using a densitometric video camera (Image Technology Methods) and a µVax II computer (Digital Equipment Corporation). The linearity of the digitization camera was confirmed (between 0.1 and 1.2 O.D. units) prior to image acquisition. Software was developed to allow the operator to collect 64 x 64 pixel images of individual molecules via an interactive selection process using the “live” 512 x 512 pixel
<table>
<thead>
<tr>
<th>Micrograph Number</th>
<th>Number of Particles Collected</th>
<th>Number of Background Regions Collected</th>
<th>Signal to Noise Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>809</td>
<td>127</td>
<td>33</td>
<td>5.35</td>
</tr>
<tr>
<td>811 (S3)</td>
<td>318</td>
<td>32</td>
<td>5.8</td>
</tr>
<tr>
<td>817</td>
<td>116</td>
<td>0</td>
<td>N/C</td>
</tr>
<tr>
<td>824 (S2)</td>
<td>180</td>
<td>32</td>
<td>6.6</td>
</tr>
<tr>
<td>858 (S1)</td>
<td>300</td>
<td>0</td>
<td>N/C</td>
</tr>
</tbody>
</table>

Table 1. Details of the numbers of particles collected from different micrographs. The signal to noise ratio is also given for micrographs in which background regions were collected. The image series designation, S1 etc. indicates the data sets used in 3D reconstruction. Image series S5 includes all the above images. (See Table 2.)

The density of particles on the micrograph was such that, at suitable digitization sizes, typically less than 8 particles would be contained within a single 512 x 512 pixel area. The low density of particles was one of the primary reasons for not adopting the use of automated particle analysis routines such as that described by Andrews. Selection of particles was carried out prior to the digitization step using the dissecting microscope to identify particles which were then marked to permit easy recognition when displayed on the computer monitor during digitization. The visual selection criteria were based on the size and intensity of the particles. Although such a visual selection criteria is initially very subjective, the final choice of the particles can be made on the basis of objective criteria such as similarity in the...
integrated intensity of the particle. The use of visual selection criteria to identify putative molecules remains the most commonplace amongst other groups working in similar fields, primarily due to the impracticality of implementing automated schemes. In all, 1041 particles were selected from the five micrographs (see Table 1). Images of background regions were collected from three of the micrographs. Signal to noise ratios were defined as, the mean pixel value in the central portion of the image (signal) minus the mean pixel value in the corner of the image (background), divided by the standard deviation of the pixels in the corner regions.

A schematic of the total reconstruction process is shown in Fig. 1. The main procedures are shown along with the required input data that must be supplied and the intermediate data sets that are produced at each step.

5.2.2 Extraction and 2D Rotational Alignment of Images

As described in Chapter 1, it is necessary to align images rotationally before they can be submitted to multivariate statistical analysis techniques. Reference-free alignment techniques (see discussion in Chapter 1 and detailed description of implementation in Chapter 4) were used to perform this alignment. Rotational alignment procedures are sensitive to intensity gradients in the background of images; indeed the rotational alignments can be dominated by the gradient rather than by the structure of the particle upon the gradient. One approach to the removal of background gradients is the use of a high-pass Fourier filter. Such filters remove the low frequency components in the image which are potentially responsible for the gradient. Because the particle in the images considered here occupies roughly half the image, its large scale variations, in terms of their spatial frequencies, will not be significantly different from those in the background, and the potential exists that these frequencies may be lost as a result of high-pass filtering. For this reason we chose not to use such a technique. Instead, the average intensities of each of the corners (9 x 9 pixels) of the image were calculated and these intensities were used in a bilinear interpolation scheme to remove local background intensity gradients from the image. The program also provides the option of setting the intensities of the corners to their global mean or setting them to a constant value.
Figure 1. Schematic showing the sequence of procedures that were followed to generate 3D density distributions from electron microscope images. The procedures are discussed in detail in section 2.

The latter option was used to correct for intensity gradients over the entire population of images in a single micrograph and to correct for slightly different exposures in the different micrographs.

Observation of the results when reference-free alignment techniques were applied to the “flattened” images indicated that the procedure was not performing as one might expect.
Images that were interpreted visually as containing similar views of the molecule were found to not align in similar orientations. Inspection of the images suggested that the local background structures were probably influencing the rotational alignment. To overcome this problem the particle was extracted from the background region.

The obvious choice for such an extraction was the use of a contouring program in which only areas of the image containing intensities above a given level would be used for alignment. Because of the high level of noise in electron micrographs and the presence of support film structure, contouring approaches often resulted in the isolation of a number of separate areas in the image which are above the threshold.

The extraction method adopted in the experiments described below was as follows. It was assumed that the total intensity in the image consists of a contribution from the background regions of the image $B$, and an additional intensity arising from the molecular signal $S$. The total intensity of the image was thus $B+S$. In the algorithm, the central portion of the image was examined and the highest intensity pixel was selected. The intensity of this pixel was recorded as belonging to the particle. The pixels that were contiguous with the selected pixel (8 of them) are then examined and the highest amongst them selected as also belonging to the particle and the total intensity of the selected pixels was incremented. The algorithm then proceeded in an analogous manner, selecting the pixel with the highest intensity amongst the pixels which were contiguous to those already selected until the intensity in the selected area was a predetermined value $S$. In this manner a single area of the image was “grown” and selected to represent the particle in the image.

To estimate the cutoff level $S$ a cumulative grey-level histogram from 1028 of the extracted particles was plotted (Fig. 2). The distribution was seen to be asymmetrical and is interpreted as resulting from the summation of two separate distributions, that of the background and that of the signal regions. Following the reasoning and procedures described by Andrews et al. the upper portion of the background region was approximated by a reflection of the portion of the curve below the modal value. The contribution of the signal to the curve was estimated from the subtraction of the two distributions. The total intensity in the signal portion of the curve accounted for 43% of the intensity in the image. However, when the images are grown to an $S$ of 43% of the total intensity of each image, the boundary...
Figure 2. Grey-level histograms derived from 1028 images from all five digitized micrographs. The "Total" and "Cumulative" histograms are normalized to 1 at their respective maxima. The "Signal" and "Background" lines are determined as described in the text.

of the particle extended far beyond that which would have been determined visually. This can be understood in light of the large overlap of background intensities and image intensities which cannot be separated. To determine a more realistic threshold, images were grown to a range of threshold values $S$ ranging from 35 to 10%. The level that best represented the individual particles was determined visually by 4 observers using images from all 5 micrographs. All observers selected the 12.5% threshold as the one that best represented the particle boundary. This threshold was subsequently used to isolate particles in all the images.

The "grown" images were subjected to the reference-free alignment procedure and the results inspected visually to determine that particles that appeared similar had been "correctly" aligned with respect to one another. This was found to be the case.
5.2.3 Image Classification

The aligned images were then classified using the multivariate statistical analysis packages available in the IMAGIC software routines. The classification programs require a "mask" to define the region of interest for the analysis. This requirement was developed to prevent structure in peripheral regions of the images influencing the classification. The masking procedure also reduces the computation time. The mask was created by interactive contouring of an image which was formed by the summation of all the aligned images. The IMAGIC programs converted the contour into a binary mask that defined the region to be considered in each image.

The number of eigenimages and other classification parameters were determined from simulations such as those described in Chapter 4 and from examination of results using small data sets of real electron micrographs. From each data set, 10 eigenimages were calculated. Classification was based upon a modulus distance measure. The modified hierarchical classification scheme of van Heel was used to perform the classification. This classification scheme uses Wards criterion of "minimum added intra-class variance" to determine the optimum partition of the images. The algorithm is modified to permit the migration of images to more similar classes following the initial classification (see Chapter 1, section 1.2.2.4). The numbers of classes were decided on the basis of giving an average of 5-10 images per class and from examination of the classification dendrograms (see Results section). The images representing the 15% of images of each class contributing most to the intra-class variance were not included in the class averages.

5.2.4 Relative 3D Orientation Determination

Following classification, images were converted from IMAGIC format to that used by the alignment programs. The quaternion based alignment determination (IQAD) programs were then applied in a manner identical to that used for the simulations in the previous chapter. Sinograms were generated from the projection images which were masked with a circle centred at the centre of mass of the image. This masking had to be done with care. If the
mask was simply an approximation to the circle defined by complete pixels, then differences in the diameter of the mask led to artifacts in the sinograms. This problem was overcome by describing the perimeter of the circle using polygons, the vertices of which were defined by the points at which the circle intersected pixels boundaries. The number of check projections for the IQAD procedure was set to 5 for each analysis and 64 iterations were used to refine the alignment. After the basic IQAD alignment had been performed, the mIQAD procedure was used to refine the alignment by rejecting outlying data points in the manner described in Chapter 4.

When the orientation of the projections had been determined, the 3D density distributions were reconstructed using the filtered back-projection procedure described by Harauz.26 To facilitate visualization of the data sets, the resulting 3D density information was reformatted. The apE software system (Ohio Supercomputing Center)27 was used to allow visualization of the density as a solid object. The Chain molecular modelling software28 was used to display the data as a wire frame pseudo electron density map following its conversion into the X-PLOR data format.29 The use of the Chain software package permitted comparison between electron-microscopy-derived volumes with the α-carbon-backbone coordinates derived from the X-ray crystallography structure as well as with the surface representation of the Klenow Fragment kindly supplied by the group of Steitz. Both visualization packages permitted the density distributions to be viewed at a user defined density threshold.

To assess the consistency of the results of the methods described above, particularly the automated orientation-determination procedures, the data was divided into three groups and individual classification, alignments and reconstructions were performed. The three data sets were formed from three of the original five micrographs that were selected for digitization. A fourth reconstruction was performed using all the images from all five micrographs.

5.3 Results

Although four data sets were described in the methods section, it is sufficient to report in detail the results of just two of them. Results from work with 300 images from a single
Figure 3. An example of an area of the electron micrograph 858 (corresponding to data set S1). The individual molecules of the Klenow fragment are seen to be well separated from each other and are discernable from the structure of the carbon film background. The bar in the image represents 100 nm.

micrograph plate number 858 and the complete set of 1041 images from all the 5 plates are described. These two data sets are referred to as the S1 and S5 series data sets respectively. The other two sets of data that were processed consisted of 180 images from plate 824 and 318 images from plate 811. These two sets of data are referred to as S2 and S3 respectively. An example of a typical area of electron micrograph 858 is shown in Fig. 3. The figure illustrates that the Klenow Fragment was readily discernible from the background support film. The mottle structure of the support film is also clearly seen in the figure.

Figure 4 illustrates examples of the digitized images derived from the S1 series data. Also shown in the figure are the same images after the particles had been selected from the
Figure 4. Images from the S1 data series. Row (A): original images as digitized from the electron image plate. Row (B): the same images after the contiguous region representing the particle has been selected using a 12.5% growth level (see section 2). Row (C): the same images following application of the reference-free alignment procedure. (D) the composite image that forms the reference for the reference-free alignment technique.

image using a threshold of 12.5% as described in the previous section. It can be seen from the figure that the procedure was effective, extracting the portion of the image corresponding to the particle and has not included any high intensity areas in the periphery of the image. The same group of images are also shown in the orientations they adopt following the application of the reference-free alignment procedure. It is seen from the figure that the particles have aligned in approximately the same orientation as required for the subsequent
Figure 5. Rotation angles assigned by the reference-free alignment procedure. (a) the S1 series data, (b) the S5 series data. Transitions between different micrographs in (b) occur at images 128, 447, 562 and 742. The angles are plotted as a function of image number to reveal any trends in the alignment procedure, e.g., due to sequential rotation of the reference image. No such trends are seen.

application of classification techniques. The composite image to which the images were aligned in the reference-free alignment technique is also illustrated. The composite image displays a considerable amount of structure despite the fact that it was derived from the summation of 300 images. One of the concerns with the reference-free alignment technique was that the reference would become indistinct and therefore would not function well as a reference. However, as illustrated, this is not the case. Reinforcement of the aligned images yields a suitable reference image when applied to images of the Klenow Fragment.

In Fig. 5 the rotation angles for all the images in the S1 and S5 series are plotted. The angles are plotted as a function of image number rather than as a frequency histogram to permit recognition of any alignment trends that might have occurred as a result of the reference-free alignment technique, e.g., as a result of gradual rotation of features in the composite reference image. From Fig. 5 it can be seen that such problems are not found. There is a distinct structure to the distribution of the angles in both series. In the S1 series
it is seen that the majority of images are rotated by approximately 45° or by -135°, whilst relatively few images are rotated by 135° and by -45°. Examination of the distribution from the S5 data set shows that similar patterns exist but that the preferred orientation varies from image plate to image plate. Similar distributions of angles arising from the reference-free alignment technique were seen in the other data sets. This characteristic will be discussed later.

Following multivariate statistical analysis, the positions of the images were plotted in the lower dimensional “eigenimage” space and the distributions were examined visually to determine if there was any obvious clustering of the data which would suggest the existence of preferred particle orientations. No such structure was seen in the two dimensional plots - the images forming continuous distributions.

Hierarchical classification was performed as described above. Dendrograms are used to illustrate the class structure of a data set, as in Fig. 6. For an understanding of this representation of the classification procedure it is necessary to recall from Chapter 1 that the total variance of a set of data points that has been subjected to a classification procedure is equal to the sum of all the intra-class variances plus the total inter-class variances. The hierarchical classification algorithm that was employed in this study used the criterion of “minimum added intra-class variance” to determine the most similar images which were merged into a class, or classes of images which were then merged to form a single class.

In Fig. 6 the original images are represented by points distributed across the bottom of the “root shaped” structure in each figure. Because of the large numbers of images in the original data sets, the individual images are not distinguishable. The original \( N \) images are considered as \( N \) classes. As two classes are merged to form a single class the intra-class variance of the class that is formed will be larger than the total intra-class variance of the two original classes. The merging of two classes is represented in the figure by two vertical lines extending up from each class. The length of the line is proportional to the increase in the intra-class variance resulting from the merging of the two classes. A horizontal line is used to join the upper end of the vertical lines to indicate that a single class has been formed. In Fig. 6 the single merged class is considered to lie at the right-hand end of each horizontal line. Using this representation merging classes of similar images will result in short vertical
Figure 6. Classification trees (dendrograms) arising (a) from the S1 data series and (b) the S5 data series. The vertical distance in the plot indicates the increase in intra-class variance on merging 2 classes. The original images are distributed uniformly across the base of each dendrogram in an order determined by the final classification.
Figure 7. Examples of the results of the multivariate statistical classification. Three classes of images from the S1 data series are shown. The four images in row (A) and five images in row (C) are representative of classes with a low intra-class variance, images (B) and (D) are their corresponding class average images. The images in row (E) form the class with the highest intra-class variance, image (F) is the corresponding class average.

lines whilst merging very different images represented by large verticals. A data set that contained a number of well defined classes of similar images would be recognizable by the transition from short vertical lines, resulting from the merging of similar images within a class, to long vertical lines which result from the merging of different classes of images.

The dendrograms that result from the S1 and S5 series are shown in Fig. 6. The structure of the two plots is relatively similar. In neither case is there evidence for the
Figure 8. An example of the form of the sinograms and sinogram correlation functions that result from the S1 data series. The two images giving rise to the sinograms are labelled (A) and (B). Sinogram (C) (plotted vertically) corresponds to image (A), sinogram (D) (plotted horizontally) to image (B). The sinogram correlation function (E) is plotted as a grey-level intensity distribution, bright areas corresponding to high correlation. The grey-level mapping changes at the 0.9 correlation level to permit visualization of structure in the high correlation region. The position of the maximum correlation is marked with a cross.

existence of thresholds that would suggest a few clearly defined classes. It should be noted that these dendrograms are illustrative of the classification that is determined using a hierarchical technique and Wards criterion for merging of neighbouring classes. The migrations of images to different but more similar classes that occurs during the modified hierarchical classification procedure implemented by the IMAGIC software routines cannot be illustrated easily in such a diagram. However, the number of images that were migrated
between classes was small, 7.8% and 23.8% in the S1 and S5 series respectively. Because no obvious partition level was indicated by the dendrograms, the number of classes was decided on the basis of the potential signal-to-noise-ratio enhancement that would be derived from summation of the images and from the desire not to merge classes representing different projection directions. The S1 series was classified into 64 classes, whilst the S5 data was partitioned into 100 classes.

In Fig. 7, examples of classes with high and low intra-class variance derived from the S1 series are illustrated. In each case, the class-average image that results from summation of the class members is also shown. It is seen that the images within the classes do resemble one another. The signal-to-noise ratio enhancement that results from the summation of the images cannot be calculated from this data set, as the background of the image was removed during the isolation of the particle for the reference-free alignment procedure. However, it is seen that the class average images do represent a consensus view of the class of images of
Figure 10. Projection directions assigned by the IQAD procedure to (a), the S1 series data and (b), the S5 series data. The solid symbols indicate that the projection direction has a value of $\beta > 90^\circ$; the projection direction is indicated in its equivalent position of $\alpha + 180^\circ$, $180^\circ - \beta$.  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the particle as one would expect. Other classes from the other series were seen to show similar degrees of similarity.

Figure 8 shows an example of two sinograms and their sinogram correlation function. The sinogram correlation function is seen to have very broad maxima, which partly explains why determination of projection directions on the basis of common line projections is sensitive to noise in the projections. Noise in the line projections can easily cause a large shift in the position of the maxima of the sinogram correlation function.

In Fig. 9 the positions of the maxima of the sinogram correlation function are plotted for all the 415 common axes determined in the S1 series and for the 667 maxima determined in the S5 series. Both plots indicate a clustering of the maxima in 4 distinct regions. Similar clusters were seen in the other data sets. The phi 1 and phi 2 axes refer to the angle at which the line projections were found to correlate maximally.

Figure 10 illustrates the projection directions that were determined using the IQAD
Table 2. Details of the alignments determined by both the IQAD and mIQAD procedures. The Number of common axes refers to the number of axes used for the final alignment determination.

<table>
<thead>
<tr>
<th>Image Series</th>
<th>Alignment Procedure</th>
<th>Number of Images Falling in Each Quadrant</th>
<th>S.C.E.</th>
<th>Number of Common Axes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0-90°</td>
<td>91-180°</td>
<td>181-270°</td>
</tr>
<tr>
<td>S1</td>
<td>IQAD</td>
<td>7</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>mIQAD</td>
<td>6</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>S2</td>
<td>IQAD</td>
<td>7</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>mIQAD</td>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>S3</td>
<td>IQAD</td>
<td>17</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>mIQAD</td>
<td>17</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>S5</td>
<td>IQAD</td>
<td>25</td>
<td>19</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>mIQAD</td>
<td>36</td>
<td>37</td>
<td>13</td>
</tr>
</tbody>
</table>

procedure. Both of the illustrated sets of projection directions indicate that the projections are more concentrated in two of the four quadrants of the hemisphere. A similar observation could be made for the images from plate S2 series data but was not found in the S3 series. In Table 2 the numbers of projections falling in the different quadrants are tabulated. The self-consistency errors (S.C.E.) for the different alignments are also tabulated.

In Fig. 11 the distributions of the individual self-consistency errors for the S1 and S5 data are plotted. These distributions are equivalent to that plotted in Fig. 5 of Chapter 4. Examination of Fig. 11 reveals that the individual S.C.E.s do not show any dependency on their position in the alignment scheme. As shown in Table 2, the average error is
Figure 12. Projection directions assigned by the mLOAD procedure to (a), the S1 series data and (b), the S5 series data. The solid symbols indicate that the projection direction has a value of $\beta > 90^\circ$, the projection direction is indicated in its equivalent position of $\alpha + 180^\circ$, $180^\circ - \beta$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 13. The reconstructed volumes from the S1 (A and C) and S5 data series (B and D). In images A and B the volumes are displayed at an intensity threshold which corresponds to the expected volume of the Klenow fragment. Images C and D show the volumes at a threshold corresponding to 50% of the volume of the Klenow fragment.

approximately 40° in each of the plots. The distribution does not have the outlying, high individual S.C.E., points that were seen when the IQAD technique was applied to simulated data sets as described in Chapter 4. It was this distribution of outlying S.C.E. values that led to the development of the mIQAD procedure.

The mIQAD procedure was applied despite the distribution of S.C.E.s not being similar to that shown in Chapter 4. The resulting distributions of projection directions are shown in Fig. 12 and the values of S.C.E. are given in Table 2. In both examples the projections seem to concentrate further in just two quadrants of the projection direction
hemisphere. A similar effect was seen in the S2 data set and, to a lesser degree, in the S3 data set. In the S1 data set, 23% of the common axes were rejected by the mIQAD procedure, in the S5 data, 7% of the data were rejected. The concentration of projection directions resulting from the application of the mIQAD procedure was a source of concern and the mIQAD derived projection directions were not used for 3D reconstruction, (see Discussion).

Following 3D reconstruction using projection directions determined by the IQAD procedure the resulting volumes were displayed using the rendering software in the apE display package. The volumes from the S1 and S5 data sets are shown in Fig. 13. The volume corresponding to the expected volume of the Klenow Fragment and to 50% of that volume are shown. When the full volumes are examined they are seen to be rather featureless although they do possess some asymmetry and there is some indication of a small groove in the structure derived from the S1 data set. The 50% volumes are shown to try to determine if any of the cleft structure of the Klenow Fragment was visible in the centre of the reconstruction. The figure shows that the cleft structure could not be recognized although the small groove in the S1 data set remained. The other reconstructions were only similar in that they too were rather featureless. The only noticeable difference was the S2 series reconstruction which was reconstructed with only 30 projections and as a result showed some artifacts due to reconstructing the volume with so few views.

5.4 Discussion

It is apparent from the results that the alignment of electron microscope images has failed to provide reconstructed volumes that are either consistent between different data sets or that resemble in detail the Klenow Fragment as determined by X-ray crystallography. In this section, some of the possible explanations for the lack of consistency will be discussed, as will some of the patterns that are seen in the data that arises from the image processing procedures which precede the orientation determination itself.

The reference-free alignment procedure was seen (Fig. 5) to yield a distribution of rotation angles which indicated that the images were preferentially rotated by two angles (related by 180°). Figure 4 illustrated that the alignment procedure seemed to be determining
reasonable alignments for the electron microscope images (the efficacy of the method was illustrated in Chapter 4 using simulated images). Therefore the two preferred rotations must be a result of the initial data. A plausible explanation is that the particles adopted a preferred orientation in the original images; for example, the long axis of the particles all tend to align in a particular direction. Because two rotations are seen to predominate, the possibility that a background gradient in the image is influencing the alignments may be discounted, a gradient would result in a single preferred rotation.

The preferred rotations from the S5 series reference-free alignment, illustrated in Fig. 5, are seen to vary as different micrographs are included in the data set. The fact that the same angles do not predominate in different micrographs is probably due to the orientation of the sample not being the same during the different electron microscopy and digitization procedures.

The data sets that illustrate the positions of the maxima of the sinogram correlation functions (see Figure 9) showed that the maxima were clustered in four distinct regions. Figure 14 illustrates a model which explains this observation. The particle is represented by a "pear" shaped object in the model projection images. An arrow indicates the orientation of the line projection considered in each image. The images have been subjected to a reference-free alignment procedure and as a result, all their major axes are aligned. As the two images considered in the model are similar, their common axes will tend to occupy a diagonal line in the phi 1 versus phi 2 plot, the direction of the diagonal depends on whether the particles are parallel or anti-parallel after alignment. The off-diagonal elements will be less populated due to the alignment procedure. When the similar distributions were plotted for the simulated image data in Chapter 4, the maxima were also found to occur primarily on the diagonal.

Figure 9 shows that the diagonal is not occupied uniformly. There are two clusters of orientations per diagonal. This structure may be explained by the particle rotating about its long axis. Common lines corresponding to line projections parallel to the shorter axis will therefore be unlikely and the result will be the clustered diagonal structure seen in Fig. 9.

The distribution of errors illustrated in Fig. 11 indicates that a highly self-consistent alignment has not been achieved. The large errors may also be related to the factors used to
Figure 14. Schematic plot of sinogram correlation function maxima to explain the similar distributions seen in Fig. 8. The number in each image denotes whether the line projection corresponds to angle \( \phi_1 \) or \( \phi_2 \). The arrows in the images indicate the direction, perpendicular to which, the line projections are calculated. The structures in the images represent elongated molecules that have been aligned along their long axes. The molecule adopts orientations which are a rotation about the long axis which is parallel to the support film.

explain the structures in the reference-free alignment and sinogram correlation function data sets. If the projections in the original data set are indeed a collection of views which are related by a rotation about a single axis perpendicular to the imaging beam then the projection directions (as illustrated in Fig. 10) should describe an arc across the surface of the hemisphere. If the axis of rotation is less rigidly defined, the projection directions will tend to spread out somewhat and will eventually occupy a broad band across the hemisphere. In three out of the four data sets examined, evidence was seen for this sort of clustering of
projection directions.

The overall $S.C.E.$ for the alignments (Table 2) are extremely large when compared with those resulting from the simulation data in Chapter 4. At the level of error found in the S1 and S5 data sets a very low quality reconstruction would be expected.

The mIQAD procedure was applied to the data set to determine if it significantly improved the quality of the alignment. This was not found to be the case. It should be noted though that the distribution of individual $S.C.E.$s shown in Fig. 11 would not have suggested the use of rejection procedure that the mIQAD algorithm performs (in the following chapter, other alternative techniques for refining the alignment will be discussed). The mIQAD procedure was applied, however, and it was seen to concentrate the projection directions even more. This concentration has been interpreted as resulting from the algorithm in rejecting common axes which would orient the projections at large angles to one another. If these common axes are rejected then the projections become co-planar and overall $S.C.E.$ will be reduced.

One problem with the common axis method is that it cannot determine the relative orientations of projections which are related by rotation about a single axis perpendicular to the direction of the imaging beam. As a consequence of this, if a situation arises in which the majority of views are rotations about a single axis then the assignment of orientations about that axis will tend to contain large errors. In explaining the structure of the data in Fig. 9 it was necessary to invoke just such a collection of projections. This might account for the large values of the $S.C.E.$ resulting from the orientation determination.

5.5 Conclusions

In this chapter we have described our investigation of the 3D structure of the Klenow Fragment using dark-field electron microscopic studies. The experiments described represent the first application of a common axis based 3D alignment algorithm to real dark-field electron micrographs. As was stated in the Introduction it was recognized that, due to instrumental and technical limitations, the dark-field micrographs were probably not optimum in terms of the radiation damage and specimen imaging conditions. However, preliminary
observations (see Chapter 1, Fig. 2) indicated that the quality of the images obtained was sufficient for us to be able to recognize major structural features of the Klenow Fragment that had been described previously from X-ray crystallography studies.

In an attempt to improve the image quality, classification techniques that were originally developed for application to the study of macromolecular complexes using bright-field electron microscopy of stained specimens have been applied to dark-field images of the Klenow Fragment. Inspection of the classes of images that were determined revealed that they did indeed share similar features.

A reference-free alignment technique was developed to permit determination of the relative orientation of the images in 2D prior to classification. Bringing similar images into register with one another is a prerequisite for the application of automated image classification techniques. Preliminary work using molecule images containing regions of background, indicated that the rotational alignments were influenced strongly by features in the background region. To address this concern a method has been developed to allow the extraction of the particle from the image so that alignments are based on the particle alone and are not influenced by the background. Following this modification, inspection of the aligned images indicated that similar images had been brought into register with one another. However, a small percentage of the images were rotated to positions that would not have been expected from visual examination of the images.

Multivariate statistical analysis techniques have been implemented to determine classes of similar images. Examination of the selected classes indicated that the method performed as required. Once again, it was noted that the automatic collection algorithm sometimes resulted in classes of images that, from visual examination, would not have been identified as belonging to the same class.

Interpretation of the results of applying automated alignment and classification techniques to real image data sets is rather complicated. Knowledge of either the "correct" rotational alignments or the "true" class structure of the data is not available. However, there are a number of options that allow the validity of such algorithms to be tested. These options generally test the consistency of the results obtained when the data set is reordered or partitioned into smaller groups. We have demonstrated the stability of the reference-free
alignment technique in this manner: the sequence of simulated images presented to the algorithm was reordered and it was verified that the relative rotations applied to each image remained unchanged. Other workers in the field have been able to test the validity of classification techniques by partitioning the classes into two separate subclasses and then determining the resolution between the subclasses. This technique is only possible, however, when a large number of images occupy similar orientations. Inspection of the class structure of the Klenow Fragment data sets suggested that such large classes were not present, and so this sort of analysis was not pursued at this time.

The application of the IQAD procedure to the determination of projection directions from dark-field electron microscopy images of the Klenow Fragment does not yield 3D reconstructions that show convincing similarities to the structure of the molecule derived using X-ray crystallography. The experimental results are consistent with the majority of particles occupying a limited range of orientations with respect to the imaging beam. It is noted that this range of orientations, rotations about axes roughly parallel to a single axis of the molecule, would be particularly difficult for the common axis method to solve.
References


Previous chapters described the results of applying an iterative image alignment procedure based on the common axis theorem to simulated and dark-field electron microscope images of the Klenow Fragment. The alignment determination procedure worked well when applied to simulated images. Application of the same procedure to electron microscope images resulted in alignment that exhibited a much higher degree of internal inconsistency. In Chapter 5 some of the possible causes of the discrepancies between these two results were touched upon; they are enlarged upon in this chapter. Simulations to indicate the relationship between the S.C.E. and 3D reconstruction quality are described. This simulation indicates that reconstructions with an S.C.E. of greater than 10° are unlikely to provide useful structural information. It is demonstrated that the quality of the 3D reconstruction may be improved by using a set of images which are selected visually as representing exceptionally high quality views of the molecule. Methods for improving the quality of the alignments in the context of improvements in the input image quality and possible modifications that might be made to the alignment algorithm are discussed. Projections are made to indicate which of these developments are likely to be most beneficial with respect to improving the performance of the orientation determination procedure. A parameter for characterizing the quality of the input image quality is discussed. Finally a new method for assessing the reliability of the projection directions determined by the algorithm is described.
6.1 Introduction

In previous chapters, the development and implementation of an algorithm that determines the relative orientations of a series of projection images has been described. Chapters 3 and 4 detailed the application of the algorithm to simulated data sets whilst Chapter 5 described the application of the technique to dark-field electron micrographs of the Klenow Fragment of DNA Polymerase I. The simulations were designed to assess the efficacy of the algorithm when presented with data similar to that expected from dark-field imaging experiments. These simulations suggested that the algorithm was robust enough to cope with such data. However, when the algorithm was applied to the problem of determining the 3D structure of the Klenow Fragment the results were somewhat disappointing. In this chapter, the relationships between the results of the previous three chapters will be discussed. In section 6.2, specific results derived from the studies of simulated images and electron micrographs will be analyzed and compared. A study that provides a visual illustration of the relationship between the S.C.E. and the quality of the 3D reconstruction is described. This study permits a reassessment of the reconstructions that were presented in Chapter 5. In section 6.3, methods for further enhancing the input data available to the algorithms will be discussed. These methods include different experimental techniques and modifications to some of the image processing procedures described previously. Visual selection criteria are used to select a series of high quality images which are then submitted to the orientation procedure described previously. The results of this investigation and its implications for future development of the orientation procedures are discussed in section 6.3. Further modifications of the IQAD procedure are considered in section 6.4. A new parameter to assess the performance of the alignment algorithm is described in section 6.5. Also in that section, a parameter is discussed which quantifies the quality of the input images by considering data directly related to the alignment technique.

6.2 Simulation and Experimentation

In Chapter 3 a method was presented that facilitated the determination of the relative
orientations of a large number of projection images. The method was developed to permit the 3D reconstruction of molecular density distributions from the 2D density projection images that result from dark-field electron microscopy techniques.

The input data for the method consist of the angular positions of common axes\(^1\) between projections. The method was developed and discussed in Chapter 3, using common axes positions that were determined analytically from a set of projection directions defined by their Euler angles. The alternative to determining the projection directions analytically was to construct simulated projection images and to calculate sinogram correlation functions which reveal the positions of the common axes.\(^3\) The determination of the positions of common axes from projection images requires a considerable computational effort, whereas they may be calculated analytically from a set of projection directions extremely quickly. As the method was still under development at this stage, analytical determination was much more convenient. Also, the use of analytically derived common axes positions avoided possible complications related to the accuracy of the experimental determination of these positions and the effects of the degree of asymmetry of the object under study.

Figure 1 of Chapter 3 illustrated the relationship between noise in the simulated projection images and errors in the position of the common axis, as determined by the sinogram correlation function using the 3D density distribution of the Klenow fragment as a model. Using this information, similar errors were added to the analytically determined common axis positions to determine the effect on the quality of the 3D alignment. The results of this investigation are illustrated in Fig. 7 of Chapter 3. It was determined that the method worked correctly when the positions of the common axes contained no errors (equivalent to the situation of the projection images containing no noise). When errors were introduced in the common axes positions equivalent to those that would be expected from images with a signal to noise ratio of 5.0, it was seen that the alignments were determined such that they had a self-consistency error (S.C.E.) of 20° and a residual system error (R.S.E.) of just 9°. In other words, the relative positions of the x- and z-axes of the projection coordinate systems have an average positional error of 9°.

From the work of Harauz, it was expected that an Euler angle error of 10° would reduce the quality of the 3D reconstruction, as measured by a 3D correlation coefficient, only
slightly, $= 15\%$. At such an error level, the structure under study, a model of the DNA super helix in the nucleosome, was still easily recognizable in the 3D reconstruction. This observation suggested that the alignment algorithm discussed in this thesis would provide sufficiently accurate angular information to permit useful 3D reconstructions. As a result, studies were undertaken with simulated projection images and sinogram correlation functions to determine the positions of the common axes. These simulations, which included variables such as image noise and orientational variations within classes of images, were described in detail in Chapter 4.

The results of these studies indicated that the basic alignment technique (IQAD), and an adaptation of it (mIQAD) were able to achieve satisfactory alignments of groups of simulated projection images of the Klenow Fragment at a signal-to-noise ratio of 3.1 even when the projection directions of the images in each class were perturbed over a $20^\circ$ range in $\alpha$ and $\gamma$, and by $10^\circ$ in $\beta$. The most adverse conditions simulated yielded an R.S.E. of just $7^\circ$. The image simulations described in Chapter 4 incorporated the image processing techniques of reference-free alignment and multivariate statistical classification. These processing steps were included in the simulations because it was expected that they would be required in subsequent experiments using electron micrographs. Although these techniques introduced a small number of misregistration and misclassification errors, neither of these errors had serious consequences on the determination of relative projection directions. The results of these simulations suggested that the alignment technique would be successful when applied to dark-field electron micrographs of the Klenow Fragment which had previously been determined to have a signal to noise ratio of approximately 6.0.

The application of the IQAD and mIQAD procedures to these images was described in Chapter 5. The only procedural difference between the simulations in Chapter 4 and the work with the electron micrographs was the isolation of the particles from their backgrounds when the electron microscope images were considered. This step was necessary because the micrographs contained some features in their background regions (probably due to structure in the carbon support film) which adversely affected the reference-free alignment process.

In Chapter 5 the 3D reconstructions that resulted from two sets of electron microscopy data were illustrated and were shown to be rather featureless (see Fig. 12, Chapter 5). The
application of the IQAD procedure to four sets of real images resulted in alignments that had self-consistency errors on the order of 40-50° (see Table 2, Chapter 5). The use of the mIQAD procedure produced alignments with lower self-consistency errors but concerns about the resulting projection directions and the suitability of the input data indicated problems with this technique (see section 6.4 below). When dealing with real images, the only internal measure of the quality of the alignment available is the S.C.E. However, it is difficult to interpret the effect of a given S.C.E. on the quality of the 3D reconstruction. To examine the relationship between the S.C.E. and quality of a corresponding 3D reconstruction the simulation described in section 6.2.1 was performed.

In section 4.3.4 it was noted that the addition of Gaussian noise to the simulated images of the Klenow Fragment will not accurately represent the degradation of the image of the molecule that occurs in the real situation as a result of such processes as radiation damage and deformation. Visual comparison of the simulated images and the real images, e.g. Chapter 4, Fig. 1 and Chapter 5, Fig. 5, reveals that the noise in the two cases is also somewhat different. There is seen to be more structure in the background of the real images indicating short range spatial correlation of the noise, probably as a result of structure in the carbon film. The effect of this difference on the validity of the intercomparisons between simulations and real images based on signal to noise ratio is extremely difficult to predict. The spatial correlation present in the real noise probably affects the orientation determination process in a detrimental manner. One possible way to improve the validity of predictions based on simulated images would be to mathematically characterize the spatial correlation of the noise in the background of the real images. However, a much simpler approach would be to use real images of background areas to provide the underlying noise patterns in the simulated images.

6.2.1 Investigation of the Relationship Between the S.C.E and the Quality of the 3D Reconstruction

A set of 64 projection images of the X-ray crystallography derived density distribution of the Klenow Fragment were generated using a random distribution of Euler angles. The positions
of the common axes vectors between these projections were determined analytically. Error angles were added to the Euler angles, and the positions of the same common axes vectors under the new rotation were determined. Self-consistency errors, the angular distance between the pairs of common axes vectors, were calculated. Three-dimensional reconstructions were then performed using the same erroneous Euler angles. The quality of the 3D reconstruction, compared to the original X-ray density data, was assessed using a 3D correlation coefficient.\textsuperscript{4}

In Figure 1 the relationship between the error in Euler angle and the S.C.E. is plotted. As expected, the S.C.E. increases as the error in the Euler angle increases. The plot suggests

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Relationship between the S.C.E. and error Euler angles used to define projection directions. The 3D correlation coefficient between the reconstructed volume and the original data set used to create the projection images is also indicated. (Note: the reconstruction using the correct Euler angles does not correspond to a 3D correlation value of 1 due to the limited number of projections used for the reconstruction.)}
\end{figure}
Figure 2. Examples of reconstructions of the Klenow Fragment corresponding to different self-consistency errors. The reconstructed volume is in the same orientation in each of the four images. The projection direction corresponds to a view looking towards the DNA binding cleft. A: S.C.E. = 0°, B: S.C.E. = 9°, C: S.C.E. = 27° and D: S.C.E. = 44°.

an approximately linear relationship between the two errors. In Figure 2, the reconstructed volumes corresponding to a variety of S.C.E.s are illustrated. The figure depicts the Klenow Fragment oriented to show the DNA binding cleft, one of its major structural features. The volumes were thresholded such that the displayed volume corresponded to the theoretical volume of the Klenow Fragment. The 3D correlation coefficient for these volumes is plotted in Fig. 1 as a function of Euler angle error. The 3D correlation coefficient decreases in an approximately linear fashion as the Euler angle error is increased. This relationship was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
expected from the previous work described by Harauz and Ottensmeyer. Examination of the 3D reconstructions shown in Fig. 2 indicates that at an S.C.E. of 9° the reconstruction is very similar to the original volume. In the reconstruction with an S.C.E. of 27° the main features of the structure are becoming obscured, whilst at an S.C.E. of 44°, the DNA binding cleft is very poorly defined.

This study provides only an example of the relationship between S.C.E. and the quality of the 3D reconstruction. There is not a simple relationship between Euler angle error and S.C.E. just as there is not a simple relationship between Euler angle error and 3D correlation coefficient. Similarly the quality of the reconstruction will change depending on the threshold considered. Structures close to the centre of the reconstructed volume are less affected by poor projection alignment than those that lie towards the periphery of the volume. However, Fig. 2 does give some visual indication of the quality of reconstruction that may be expected at different levels of S.C.E.

The reconstructions that were obtained from the electron microscopy data sets may be reassessed in light of this study. Reconstructions from the S1 and S5 data sets have S.C.E. values of 47.3° and 46.7° respectively. The results presented in Fig. 2 suggest that, even if the electron microscope had provided ideal projection images which were aligned with an S.C.E. of this order, the structure of the Klenow Fragment would not be easily recognizable.

The implications of the study presented in Figs. 1 and 2 suggest that it is not possible to draw useful conclusions about a reconstructed volume unless the S.C.E. of the alignment is around 10°. For this reason, the reconstruction of the Klenow Fragment depicted in Chapter 5 was not claimed to be an accurate representation of the structure of the molecule.

The representation of the reconstructions used in Fig. 2 suggests that there is a sharp boundary to the molecule. It should be remembered that the display illustrates a single density threshold and that the 3D data set is actually a gradually varying, diffuse set of intensities.

6.2.2 Verification of the Reliability of 3D Structural Determinations

The experimental design of the work discussed in Chapter 5 was such that three independent reconstructions from images from three different micrographs would be produced and
compared to allow the success of the IQAD method to be assessed. A number of 3D resolution measures, other than the 3D cross-correlation coefficient, that permit comparison of independent reconstructions have been employed by other groups. While comparison of two aligned 3D data sets is relatively easy, a more difficult problem is the preliminary task of aligning the two data sets with respect to one another. One method for determining such an optimum alignment is a simple rotational (and possibly translational) search. A second method proposed by Carazo and Frank employs autocorrelation functions of the volumes to determine an optimum alignment. A consideration for both of these procedures is the choice of threshold at which the volumes should be compared. Different thresholds will not necessarily result in the same relative orientation of the volumes.

It was decided not to perform a quantitative intercomparison of the 3D reconstructions that resulted from the use of the three different data sets. There were 1 number of reasons for this decision. Firstly, the S.C.E. values of the alignments were high, suggesting that the algorithm had not been able to determine the relative projection directions accurately, and as a result one would not expect similar reconstructions (see section 6.2.1). Secondly, visual inspection of the reconstructions revealed that the structures were in fact only similar in that at large thresholds they were rather featureless. Finally, on a more practical level, it was decided that subsequent development work would be more profitably directed at improving the accuracy of the alignment rather than in quantifying the degree of similarity of the reconstructions that were visually different.

### 6.2.3 Image Processing Procedures

In Chapter 5, a possible explanation for the poor performance of the algorithm when applied to electron microscopy data was described. The explanation was based on the hypothesis that the molecules did not lie at random orientations to the carbon film but were preferentially distributed such that a long axis of the molecule lay parallel to the support film. Different views of the molecule were interpreted as resulting from rotations about this axis. This explanation was supported by data produced by the reference-free alignment procedure and the distribution of the maxima of the sinogram correlation functions. The algorithm will work...
best if it is presented with projections at large angles to one another. Therefore, the preponderance of views related by a rotation about (roughly) a single axis, goes part way to explaining the high S.C.E. values. However, there are other factors that should be considered. In this section, the implications of the choice and application of some of the image processing techniques will be discussed.

It should first be noted the simulations described in this thesis, particularly those in Chapter 4, were intended to be as realistic as possible. In this manner, each of the individual image processing steps, reference-free alignment, classification techniques, sinogram generation etc., could be rigorously tested under well-defined conditions. However, there were differences between the characteristics of and the procedures applied to simulated and real dark-field images. The most significant difference in procedure was the particle extraction techniques that were employed prior to reference-free alignment when using data derived from electron micrographs.

The question as to whether or not the particle extraction procedure might have contributed to the high S.C.E. should be addressed. The threshold that was used to define the boundary of the particle was decided by subjective observation of the results of applying the selection process to a randomly selected subset of images using different thresholds. Although all observers selected the same threshold value there will obviously be cases when the automatically selected region of the image is not the same as that which would be determined visually. In such cases, assuming the observer were able to identify the correct threshold, the automatically selected image of the particle will incorrectly represent the projection of the mass distribution of the particle and this will lead to errors in the assignment of the positions of the common axes for this image.

An alternative procedure for isolating the contribution of the particle was discussed in Chapter 5. This method would employ an intensity contour to determine the boundary of the particle. Reservations were expressed about the use of this technique when applied to dark-field images; when the intensity threshold is set to a value close to the intensity of the background regions of the image the technique tends to isolate a number of separate regions in the image. Some of these regions are clearly associated with the background rather than the particle. Increasing the contour level allows a single region to be selected but the contour
often falls within the particle itself. One approach to overcome this problem would be to use a low-pass filtered image to generate the contour around the particle. Whilst this method may be slightly more robust than the one adopted in Chapter 5, it must be applied with care. The choice of intensity threshold must be made by the user. A rational choice of threshold might be to select a level based on the mean and variance of the background regions of the individual images. A profitable avenue for future investigation would be a comparative study of the particle extraction procedure that was described in Chapter 5 and the technique that uses contours of low-pass filtered images.

Finally, in considering the problem of particle extraction it should be remembered that this procedure is only necessary as a result of non-uniformity in the carbon support film. If these non-uniformities can be reduced then the recognition of particle boundaries will be simplified and, in the case of extremely smooth support films particle extraction, may not even be necessary. This is discussed further in section 6.3.3.

Classification techniques, which permit the recognition of similar images that may then be averaged to enhance signal-to-noise ratio, were applied to both the simulated images and to processed micrographs. It was discussed earlier that such an enhancement of the images will only occur if the images do indeed represent identical views of the molecule. One of the assumptions at the outset of this work was that a molecule such as the Klenow Fragment would exhibit random orientations with respect to the carbon support film. A consequence of this assumption is that it would be highly unlikely that two identical images of the molecule would be seen. The more probable situation would be that similar but non-identical images are recognized and averaged. The resolution of the average image would thus be limited by the degree of similarity between the images forming the sum. The summation of non-identical images will have implications for both the determination of the relative orientations and subsequent 3D reconstruction.

Essentially there are two options in performing 3D orientation determination and reconstruction from a population of images of molecules in random orientations. The first option is to perform the orientation determination using the raw images and to then generate a 3D reconstruction on the basis of those alignments. The second option is to sum similar images and to use the averaged images to perform the alignment and reconstruction. The
disadvantage of the former approach is that, because of low image signal-to-noise ratios, the
determination of the relative orientation of the projections is more likely to be in error. The
advantage of the approach is that the information in the original images is preserved for the
3D reconstruction process. If averaged images are employed, the determination of relative
projection directions will potentially be more accurate but, the 3D reconstruction will be
performed with images that have a lower resolution than the originals as a consequence of
averaging similar, but not identical images.

The decision as to which of these two approaches should be adopted is related to the
quality of the classes determined by the multivariate statistical techniques. If the images in
the classes are identical except for noise then the classes will exhibit a small intra-class
variance and the summation approach should be adopted. As intra-class variance increases,
the benefits of averaging decrease. Therefore, analysis of the clustering of classes resulting
from principal component analysis and classification of the images may indicate which
approach should be adopted.

In the work described in Chapter 5, the approach of averaging classes of images was
adopted. A number of factors influenced this decision. Firstly, direct observation of the
electron microscope images revealed a number of classes of similar views of the particle.
Secondly, the simulation work described in Chapter 3 determined that the 3D alignment
algorithms were robust enough to permit determination of the projection directions of
averaged images even when the molecules in the images were perturbed over a 20° range.
Thirdly, the error in determination of the maxima of the sinogram correlation function was
known to increase rapidly at low signal-to-noise ratio (Fig. 1, Chapter 3) indicating that low
signal-to-noise ratio images were undesirable.

To avoid creating classes that contained two independent classes of images the number
of classes was chosen to be large. One possible consequence of this decision is a loss of
signal-to-noise ratio enhancement in some of the image classes.

Another consideration in the decision as to whether to average images or not is the
reliability of the multivariate statistical classification when applied to electron microscopy
data. The above discussions describe the situation in which images are considered similar in
the sense that they result from closely related projection directions. If the classification
technique determines the images to be similar for other reasons, i.e., the eigenimages do not correspond directly to orientational variations of the molecule, then the quality of the averaged images will again be decreased. This concept has been discussed by van Heel. However, the work that was done with simulated images in Chapter 4 and the observation of classes of images derived from electron microscopy data (see Fig. 6, Chapter 5) provided
Figure 4. The distribution of the maxima of the sinogram correlation functions for the images shown in Fig. 3. Each cross on the plot represents the position of the maximum in a single sinogram correlation function.

A measure of confidence that the multivariate statistical techniques were indeed classifying images on the basis of their structural features and thus their projection directions.

A general comment might be made with respect to all the computer image processing procedures that have been implemented in this thesis. Whilst these techniques are often regarded as conferring objectivity on the analysis of images, they almost always reflect the subjective biases of their designer. Indeed, without these biases it would be very difficult to develop this type of algorithm at all. Two obvious examples of such subjective modifications have been described in this thesis. In one case, the results of the reference-free alignment procedure were analyzed and it was decided that particle extraction was required to improve the 2D alignment of the images. In the other example, the IQAD procedure was modified to
Figure 5. Projection directions assigned by the IQAD procedure to the images shown in Fig. 3.

reject outlying, putatively erroneous common axis positions.

6.3 Improving the Quality of the Input Images

6.3.1 Visual Selection of High Quality Images

In the previous section, the possibly detrimental implications of the use of particle selection and averaging techniques were discussed in relation to alignment determination and 3D reconstruction procedures. The simplest technique to avoid these problems is to use the raw images themselves. Such a reconstruction is described below. Images of the Klenow Fragment from micrograph 81 were examined and those of high quality were selected. The selected images are shown in Fig. 3. The subjective criteria used to select the images included requirements that the background in the image was smooth, that the particle boundary was
easily delineated and that the particle was both contiguous and appeared to have retained its integrity. Image selection also considered the desire to obtain a variety of different views of the object.

Sinograms were generated from the images and sinogram correlation functions were computed to determine common axis positions. The positions of the maxima of the sinogram correlation functions are illustrated in Fig. 4. As expected, the diagonal nature of the plot that was seen following alignment of the images using the reference-free alignment technique (see Fig. 8, Chapter 5) has disappeared. The distribution of maxima is also much more uniform, indicating that a wider variety of projection directions are represented by the selected images. There are however still two regions that are not well represented, indicating that certain orientations of the molecule are not seen. In Fig. 5 the projection directions, as determined by the IQAD procedure, have been plotted. The S.C.E. for this alignment was 37.6°, somewhat improved over values seen earlier in Chapter 5. It is seen that the projection directions are still not uniformly distributed. Application of the mIQAD procedure caused a concentration of projection directions similar to that described for other data sets in Chapter 5 (the S.C.E. fell to 26.7° and 12% of the data were discarded). In Fig. 6, three views of the resulting 3D reconstruction using the IQAD derived projection directions are shown. The reconstruction is illustrated at thresholds corresponding to 100%, 50% and 34% of the Klenow Fragment volume. It is seen that even at the lowest threshold, the volume possesses more structure than those illustrated earlier. At the higher thresholds, a groove is seen to develop in the structure. It could be postulated that this groove may represent the DNA binding cleft of the Klenow Fragment, although further studies will be needed to confirm this. This result is seen as encouraging.

Whilst repeating this experiment may lead to even greater confidence in the alignment techniques proposed in this thesis, the subjective image selection criteria employed need to be analyzed and replaced by objective criteria that select “high quality images” equally well.

6.3.2 Image Selection Criteria

One possible explanation for the improvement in the reconstruction when high quality images...
Figure 6. The reconstructed volume obtained from the images shown in Fig. 3. The upper row of images (A,D,G) are shown at a threshold corresponding to the theoretical volume of the Klenow fragment. The central row of images (B,E,H) are displayed at a threshold corresponding to 50% of the Klenow Fragment volume. Images C, F and I correspond to 34% of the Klenow Fragment volume. The volumes in each column are all in the same orientation. The orientation of the central column of images is related to the left column by a rotation of 90° about a horizontal axis. The images in the right-hand column are related to the central column by a rotation of 90° about a vertical axis.

were selected could be that the images represented views of molecules that had not been damaged or distorted as a result of the imaging process. If this is indeed the case then it would be desirable to develop an algorithm that could recognize the “undamaged” images directly. While there are a number of parameters that may be used to describe digitized
images,⁹,¹⁰ the recognition task proposed is extremely complex. The human operator is able to recognize images of objects in different orientations and to recognize that these images share some qualities that indicate that the object is intact. Because of the complexity of this problem, direct automatic recognition of undamaged molecules will probably not be possible for a number of years. Hopefully, workers in other image analysis disciplines will develop algorithms that can be "trained" on a series of specimen images to recognize certain qualities within a set of images.

However, there is another approach that may be adopted to recognize undamaged, high quality images. The approach assumes there is a small population of undamaged images and a larger population of more severely degraded images. If a large number of images is collected then one would assume that the images of undamaged molecules might resemble each other more closely than those of the damaged molecules. Careful examination of a classification of images similar to the multivariate classification described previously might reveal these classes of undamaged images. Two possible objective criteria that could be used to determine which classes contain the high quality images are lower than average intra-class variances and high 2D correlations between images within a class. This type of study is easy to implement and adoption of this sort of approach will probably be of great benefit to any future studies.

Another possible route for improving the quality of the input images would be to collect more images. These images could then be processed using the same techniques described in Chapter 5 but more rigorous selection criterion could be enforced. In the experiments described in Chapter 5, all the images that were originally selected were used in the reference-free alignment procedure and the principle component analysis. Only during the summation of each class of images, were the "worst" 15% of each class rejected. If a larger pool of raw images were initially available it may be possible to screen images on the basis of a smaller variation of their integrated intensity and also to reject whole classes of images which possess large intra-class variance.
A major source of noise in the dark-field images used in these studies arises from structure in the carbon support film. This support film was prepared by indirect evaporation of graphite onto mica.\textsuperscript{11} Whilst these films are some of the thinnest and smoothest currently available, it has been suggested that the use of an electron gun to evaporate the carbon may result in films with less intrinsic structure.\textsuperscript{12} The use of this technique is currently under investigation in our laboratory.

The use of a scanning transmission electron microscope (STEM) also offers the possibility of improving image quality. The STEM detects scattered electrons electronically and its collection efficiency is potentially five times greater than that of the Philips EM 300 microscope used to acquire the images used in the studies described above. The enhanced collection efficiency is primarily a result of the geometry of the electron detectors in the STEM which permits more of the scattered electrons to be detected. The enhanced collection efficiency will give a signal-to-noise ratio improvement over images collected with a similar dose in the Philips machine or alternatively will permit the dose to the sample, and thus the radiation damage, to be reduced whilst maintaining the signal-to-noise ratio. To exploit the benefits offered by the STEM, collaborative dark-field imaging experiments have begun between our laboratory and the Electron Microscope facility located in Brookhaven, New York.

In Chapter 1 the advantages of the use of cryo-microscopy techniques were discussed. Low temperatures do not reduce the number of interactions between the electron beam and the molecule but the damage caused by the interaction is not manifest because the low temperatures prevent migration of fragments resulting from the damage.\textsuperscript{13} One of the most probable causes for the inconsistency between the results of the simulations and the experiments using the electron microscope would be radiation damage of the molecule when exposed to the electron beam at room temperature. Damage processes will result in the structure in different images no longer being identical. If the effect of this damage can be reduced either through the use of lower temperatures, lower exposures or the use of chemical fixatives which preserve the integrity of the structure, then the quality of the input images
will be improved. Unfortunately, the effects of radiation damage on a molecular scale are poorly understood and thus specific model simulations of the effects of such damage on the alignment methods described in this thesis cannot be carried out.

Loss of native structure of the particle might also result from the specimen preparation procedure known as air-drying which was employed to dehydrate the sample prior to imaging. Whilst certain reservations exist about the use of critical point drying techniques (see Chapter 1, section 1.1.2) their use might have two potential benefits. Critical point drying will remove the effect of surface tension forces during dehydration of the molecule and could potentially lead to the molecule being observed in a wider variety of orientations. The benefits of a more random distribution of orientations were described earlier in relation to the results described in Chapter 5. The use of a freeze-drying technique for specimen preparation is currently being employed in our studies with the group at Brookhaven. In this technique, the sample is instantaneously frozen in its buffer. The frozen sample is then introduced into the vacuum of the microscope and the ice is then allowed to sublime. In this manner, concerns with respect to both surface tension and the less polar environment of critical point drying are addressed.

The use of heavy atom labels provides a slightly different approach to improving the input data to the alignment procedure. The heavy atom labels should be easily detectable in both the projection images and the corresponding line projections and should improve the accuracy of the determination of the maxima of the sinogram correlation functions. One such label is the undecagold cluster which may be covalently bound to free sulfhydryl groups on the protein.\textsuperscript{14,15} The label itself has a diameter of 0.82 nm and a molecular weight of 2.2 kD. This label has been used in a number of studies either bound to antibodies to permit recognition of protein subunits\textsuperscript{16} or linked directly to a particular residue in a protein which can then be recognized in a 3D reconstruction.\textsuperscript{17} In employing heavy atom labels to enhance the determination of relative orientations of different projections, a number of precautions will be necessary. Firstly, it must be shown that the label, or labels, are consistently bound to the same residues, and secondly, the bound label must be tightly bound, i.e., not free to move with respect to the protein. Both of these conditions must be satisfied because the common axis method demands that the 3D mass distributions that are projected into two dimensions.
are identical in each projection.

The Klenow Fragment has a single cysteine residue that is located on the surface of the molecule at the end of one of the α-helices which underlie the DNA binding pocket. This observation suggests that it may be extremely profitable to pursue labelling the molecule with the undecagold particle.

There are thus a number of experimental techniques that might be employed to improve the quality of the input images that are subsequently used in the 3D orientation and reconstruction procedures. The quality of the input images is undoubtably the most important factor in determining the success of the alignment and 3D reconstruction procedures. If the quality of the input images can be improved, this may obviate the need for some of the image enhancement techniques discussed earlier. As a consequence, the problems associated with particle extraction, image registration and classification will also disappear. For this reason, it is recommended that improvement of input image quality should occupy a high priority as future development of the alignment techniques are considered. All of the techniques described above, with the exception of the labelling procedures, are easily implemented given the correct apparatus.

6.4 Future Development of the Alignment Procedure

In the previous section, methods that may lead to improvements in the quality of the dark-field electron micrographs which form the input data for the 3D orientation and reconstruction problem were discussed. Input image quality was one of the possible causes of the discrepancy between the alignment results with simulated and real data. In this section possible modifications to the 3D orientation determination algorithm will be discussed. It is hoped that these modifications will lead to more accurate determination of the relative 3D orientations of the image by adapting the algorithm to use the input data more effectively. For example, examination of the individual S.C.E. spectra that resulted from the application of the orientation determination technique (see Fig. 11, Chapter 5) indicated that modifications to the IQAD algorithm other than those described in Chapter 4 might be appropriate.

187
The mIQAD algorithm was developed as a result of observing S.C.E. spectra similar to that shown in Fig. 5, Chapter 4. From such spectra, it was deduced that most of the projections were aligned correctly whilst a few projections were poorly aligned. The mIQAD procedure assumed that these incorrectly aligned projections arose as a result of the incorrect assignment of the position of the common axes between certain images. Therefore, if the majority of projections were correctly oriented with respect to one another, the erroneous common axes may be detected from their large individual S.C.E.s. From work with the simulated images it was shown that this was indeed the case and application of the mIQAD procedure gave a considerable improvement in determining the 3D orientation of the images.

Application of the mIQAD procedure to the electron micrographs also reduced the S.C.E. of the overall alignment. However, examination of the resulting projection directions indicated that the algorithm was acting simply to bring all the projections to a single orientation. Examination of the images suggested that this was not the correct solution of the problem and so it was concluded that application of the mIQAD procedure to images which produce individual S.C.E. spectra such as those from the dark-field images (see Fig. 11, Chapter 5) was inappropriate.

The difference in the performance of mIQAD method when applied to real and simulated data can be explained by the shape of the individual S.C.E. spectra. The spectra from real data (Fig. 11, Chapter 5) indicated that very few of the projections aligned by the IQAD procedure have individual S.C.E.s close to zero. Therefore, it is difficult to discern which of the projections are aligned correctly on the basis of their S.C.E.s. For real images, the likelihood of errors in the determination of the position of the common axes between projections will be increased and so the IQAD-based alignment is more likely to be in error. It was seen in Chapter 4 that a few poorly determined common axes positions were sufficient to lead to an error in the determination of the orientation of the images.

There are a number of possible modifications that might be made to the way in which the common axis data is processed. In the alignment algorithm described above, all the common axis data was considered to be of equal quality. However, the potential exists to determine which pairs of common axes represent higher quality data. The simplest way to do this is to consider the value of the maxima of the sinogram correlation functions. The value
of the sinogram correlation function indicates the degree of similarity between the line projections, thus a sinogram correlation function maximum close to 1 indicates that the projections are correctly matched along their common axis; lower values indicate less agreement with respect to the assignment of the common axis positions. With this information, the projections could be ordered such that those with high sinogram correlation function maxima are considered first in the 3D orientation determination process. Using this approach, the early assignments of projection direction should be more reliable and so the determination of the projection directions of the whole data set should also be more reliable.

Another possible approach combines some of the ideas of the mIQAD procedure with those of the above discussion. The mIQAD procedure removed data on the basis of large individual \( S.C.E. \) values. A similar approach is envisioned in which the data is removed on the basis of the size of the sinogram correlation function maximum. In this procedure, relative orientations would be determined using the IQAD procedure and the overall \( S.C.E. \) of the alignment would be calculated. The common axes corresponding to the lowest sinogram correlation function maximum would then be removed from the data set and a new set of relative orientations would be determined. If the alignment, as measured by the \( S.C.E. \), improved, another data point could be removed. This procedure would continue in a similar manner to the mIQAD procedure until no improvement in the overall alignment results from the removal of a data point.

One problem with the mIQAD procedure and the similar procedures proposed above is that the \( S.C.E. \) minima is unlikely to be a global minimum. To address this concern it may be necessary to allow the algorithms to continue to remove data following an increase in the overall \( S.C.E. \) and to continue to monitor the performance of the algorithm. A similar concern was raised in Chapter 3 with respect to the dependency of the orientation determination on the starting triplet of projections. One approach that would address both these concerns is to employ many different starting triplets of projections to start the orientation process and to examine the ensemble of relative orientations that result.

Another possible orientation refinement procedure would be to combine the IQAD method with the iterative refinement procedure that was employed by Harauz and Ottensmeyer. The iterative refinement procedure uses comparisons between forward
projections of the reconstructed 3D density distribution and the original projections to refine the orientations. When this technique was applied to the reconstruction of the nucleosome core particle, initial assignments of projection directions were made by comparing the projection images to a model structure. The IQAD procedure, or a modification of it, could be used to assign the initial projection directions.

Finally, a radically different approach to the orientation determination problem may be considered. In many respects, the problem of trying to determine the relative orientation of projections on the basis of common axis data is similar to the distance geometry problems that are posed by the data from nuclear magnetic resonance (NMR) spectroscopy experiments. The NMR data set consists of a set of distance constraints between atoms. The problem is then to determine the atomic coordinates of the molecule that satisfy these and other configurational constraints. This problem is solved using simulated annealing approaches similar to those described in Chapter 2. To use this approach to solve the 3D orientation problem, common axis vectors would be determined in the manner described previously. The positions of the ends of the common axes, in the planes of the individual images could then be determined. The image planes would be brought together at random orientations in a single 3D coordinate system. Simulated annealing procedures would then be used to attempt to satisfy the constraint that the distance between the ends of common axis vector pairs should be minimized.

Amongst the modified algorithms discussed above, the simulated annealing based approach is particularly intriguing because it would employ exactly the same data as that used by the IQAD procedure. If the two procedures converged to the same solution, as hoped, then one could be confident that the data was being processed correctly. Implementation of the algorithm using available distance geometry packages would probably not present many problems, the main task would consist of devising a method of presenting the common axis data to the distance geometry algorithm as if it were a set of distance constraints.
6.5 New Measures of Input Data Quality and Algorithm Performance

6.5.1 Assessing Input Data Quality

In the previous section an adaptation of the mIQAD procedure was discussed which employed information about the value of individual sinogram correlation function maxima. The quality of the input data to the algorithm could also be characterized globally by determining the average value of the sinogram correlation function maxima.

A resolution-based description of the input data can also be calculated. The Fourier transforms of the line projections correspond to sections in the Fourier space representation of the data. Thus, by taking the Fourier transforms of the line projections which correspond to the maximum of the sinogram correlation functions, the degree of similarity between the projections may be expressed as a function of resolution. A similar measure was discussed by Crowther who proposed using similarities between sections of the Fourier transforms to assess the degree of preservation of icosahedral symmetry in spherical viruses.

This sort of analysis will become more important as the orientation determination procedure is applied to a variety of different data sets. As experience is gained with orientation techniques, the information from such analyses will permit assessment of the potential of the data to produce high quality reconstructions before it is submitted to the orientation technique.

6.5.2 Assessing Algorithm Performance

Monitoring the performance of the alignment determination procedures when using simulated images was a simple problem due to the original projection directions being known. When dark-field electron micrographs were considered, the only measure of algorithm performance was the S.C.E. In this section, another approach to assessing the reliability of the alignment is considered.

It has been noted in X-ray crystallographic studies that to define the quality of the structure using the same $R$ value which is used to refine the structure, can lead to problems
in the validity of the final measure. In the 3D orientation determination methods described above, a similar problem exists, the projections are aligned by maximizing the scalar product between common axes vectors and the quality of the alignment is given in terms of the average angle between these vectors. The solution proposed in the X-ray crystallography work was to remove some (≈10%) of the data from the original data set. The 3D structure is then determined using the remaining data. The agreement between the data that was removed and data derived from the model structure is then assessed in terms of a measure referred to as $R_{free}^2$. This technique is equivalent to the statistical technique referred to as cross-validation.

A similar approach could be adopted using the common axis data. Common axis data corresponding to a few of the projections would be removed from the original data set. The determination of relative 3D orientations would then be performed. The final alignment would be compared with the removed data, using the common axes positions that were originally calculated between the aligned projections and the projections that were removed for cross-validation purposes. The agreement between the aligned projections and the test data set could be assessed in terms of an S.C.E. and might be referred to as an $S.C.E_{free}$. This measure could then be further refined by removing a different set of images for cross-validation purposes and repeating the process.

6.6 Conclusions

In this chapter, the discrepancies between the simulation results presented in Chapters 3 and 4 and the experimental work described in Chapter 5 have been discussed. It is thought that the primary reason for the poorer performance of the algorithm when applied to real images relates to the lower quality of the input images. A number of reasons have been proposed to explain the poorer quality of the real images. It was also demonstrated that careful selection of higher quality input images could also lead to reconstructions with a lower S.C.E. and to a reconstruction that more closely resembled the Klenow Fragment.

A number of techniques for improving the quality of the input images have been discussed in section 6.3, and methods for modifying the alignment procedure to adapt it to the sort of data that results from dark-field images have been discussed in section 6.4. Whilst
the algorithm modifications should lead to improved orientations, it is thought that future work would most profitably be directed towards improving the quality of the input data. In section 6.5, a measure for quantifying the improvement in input image quality has been described.
References


194

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


CHAPTER 7

CONCLUSIONS

The work described in this thesis has been concerned with the application of computational techniques to dark-field electron micrographs. Dark-field electron micrographs provide information about the two dimensional (2D) distribution of mass in a molecule. Through a variety of reconstruction techniques this information may be used to generate a three dimensional (3D) reconstruction of the molecule. A requirement for such a reconstruction is knowledge of the relative orientation of different images. The development and application of an algorithm that is able to determine the relative projection directions of a large number of images is central to the work described.

A general tenet in science holds that improved input data leads to more accurate predictions. In the determination of 3D molecular structures, using electron microscopy techniques, the input data is in the form of images. Maximum entropy methods were investigated in an attempt to improve the quality of the, usually noisy, images that are collected in dark-field microscopy. Whilst maximum entropy methods appeared to be directly applicable to the image enhancement problems in dark-field microscopy, the work described in this thesis revealed that application of the principles of maximum entropy alone led to artifacts in the processed images. Following this observation, a novel constraint on the autocorrelation function of the noise removed from the image was developed. This additional constraint was shown to remove the artifact in the processed image but with the penalty of increasing the computational time significantly. Following publication of this work, it was decided that application of the maximum entropy method to the image enhancement problem in dark-field electron microscopy should not be further pursued at this time.

An alternative approach to the improvement of image quality relies on the principles of the central limit theorem. It is implemented in practice by summing images that contain similar views of the molecule under study. The difficulty with the practical application of this approach in dark-field electron microscopy is related to the large number of degrees of
freedom that a molecule has when adopting an orientation with respect to its support film and consequently to the imaging beam. To exploit the principles of the central limit theorem similar views of the molecule must be determined. Principal component analysis techniques have been demonstrated to be effective in solving this recognition problem when applied to bright-field electron microscope images, the work described in this thesis represents the first application of these techniques to dark-field images.

The advantage of dark-field microscopy techniques over more conventional bright-field techniques is that heavy atom stains are not required and, as a result, the resolution of the images is potentially higher. Dark-field electron microscopy also has the advantage that there is a simple relationship between intensities in the image and the distribution of mass in the specimen. When dark-field imaging techniques are applied to the study of proteins, the molecules are generally assumed to adopt random orientations with respect to the imaging beam.

A method has been described in this thesis which permits the determination of these orientations. The method was based on the work of van Heel but makes the important contribution of extending his ideas to permit the inclusion of a large number of images. The application of quaternion mathematics to this problem proved to be most useful. Quaternion mathematics simplifies the problems posed when projection directions must be found on the basis of the maximization of a number of angular constraints. The alignment algorithm that has been described in the preceding work may be thought of as a three-step process. Firstly, the algorithm develops a scheme that determines which pairs of projections will be intercompared. Secondly, the common axes between these pairs of projections are found using sinogram correlation functions. Finally, using a combination of geometrical and iterative numerical approaches the orientations of the projections are determined.

The alignment algorithm has been tested using simulated common axis data and common axis data derived from simulated images. The results of these investigations have shown that the algorithm is capable of determining the alignment of projection images a posteriori. The results also suggest that the method should be capable of determining the relative 3D orientation of dark-field electron microscope images at the noise levels typically found in such images.
The alignment algorithm was applied to dark-field electron micrographs of the Klenow Fragment of DNA Polymerase I. The parameter that was used to measure the internal consistency of the alignments, the S.C.E., indicated that the 3D reconstructions that were derived from the micrographs were likely to be at a very low resolution. The S.C.E. of the final alignments was around 50° whereas, previous simulations in Chapters 3 and 4 indicated that a value of 10° or lower might be expected.

In Chapter 5, a possible explanation for the poor performance of the algorithm, when applied to electron micrographs, was discussed. This explanation relied on the Klenow Fragment occupying particular orientations with respect to the imaging beam. Another possible explanation for the discrepancy between the simulations and experimental results relates to the quality of the images. The dark-field images had a signal-to-noise ratio comparable to those which had been used in previous simulations; however, the signal-to-noise ratio is a poor measure of the integrity of the molecule which is extremely important to the described algorithm. It is thought that, due to the effects of specimen preparation and interactions with the electron beam, the structure of the molecule may not be equivalent in each projection.

Application of the method to images that were selected visually as being of high quality led to more encouraging results in terms of the internal consistency of the alignment, (S.C.E. of 38°), and from inspection of the 3D reconstruction. These observations and the possible improvements to the method that were discussed in the Chapter 6, indicate that the alignment procedure described above should prove valuable when applied to the problem of automatically assigning projection directions to dark-field electron micrographs.