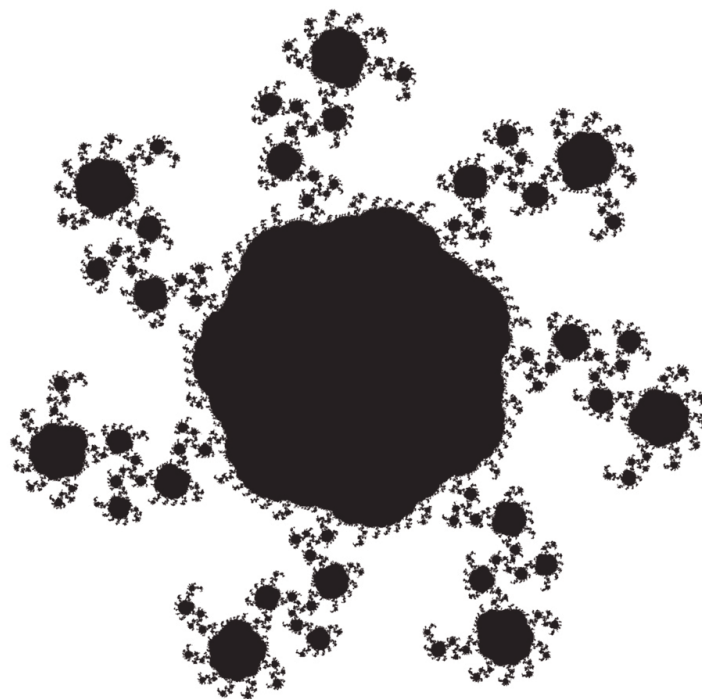


**Residual density validation
and the structure of Labyrinthopeptin A2**



Katharina Anna Christina Meindl

Göttingen 2008

**Residual density validation
and the structure of Labyrinthopeptin A2**

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August Universität zu Göttingen

vorgelegt von
Katharina Anna Christina Meindl
aus Regensburg

Göttingen 2008

D 7

Referent: Prof. George M. Sheldrick, PhD

Korreferent: Prof. Dr. Dietmar Stalke

Tag der mündlichen Prüfung: 30.10.2008

Danksagung

Ich danke

Prof. George M. Sheldrick für die langjährige Betreuung dieser Arbeit. Obwohl ich einen Grossteil dieser Arbeit in Kooperation mit einem anderen Arbeitskreis angefertigt habe, war er immer zu Diskussionen bereit und hat mich sehr unterstützt.

Prof. Dietmar Stalke dafür, dass er mich in seinem Arbeitskreis aufgenommen hat und mir in der Verwirklichung meiner Projekte alle nötigen Freiheiten gelassen hat.

ganz besonders Julian Henn dafür, dass er mit mir bei “exakt 37% Luftfeuchte” durch gute wie schlechte (und derer gab es einige) Zeiten im Laufe der Anfertigung dieser Arbeit gegangen ist. Ausserdem hat er als Starfotograf einen wichtigen Beitrag zur Bebilderung dieses Werks geliefert. Vielen Dank für deinen unerschütterlichen Optimismus und die vielen Ideen, auf denen ein Grossteil dieser Arbeit basiert!

Sebastian Merkel dafür, dass er sich immer Sorgen um mein körperliches Wohlergehen gemacht hat ☺

Tanja Tatić, weil man immer auf sie zählen kann.

Ina Dix für ihre wertvolle Hilfe bei der Integration meiner Daten und sonstige Tipps in Sachen Kristallographie.

Regine Herbst-Irmer für die gründliche Überprüfung all meiner gemessenen Strukturen und für ihre Hilfe vor allem bei Problemstrukturen.

Stephan Rühl und Tim Grüne für jegliche Hilfe, wenn Linux oder \LaTeX mal wieder nicht so wollten wie ich.

Christian Grosse, Tobias Beck, Roland Pfoh und Andrea Thorn, weil ich da oben bei ihnen immer eine Rückzugsmöglichkeit hatte.

Helmut Dehnhardt, weil er sich nicht nur mit Diffraktometern, sondern auch mit Radioweckern auskennt.

dem AK Dittrich (Birger Dittrich, Christian Hübschle und Julian Holstein), weil sie die ersten Interessenten für mein Programm waren und ausserdem für die nötige Abwechslung am Kickertisch gesorgt haben.

Burkhard Heisen für wertvolle Tipps zum “fast-and-efficient” Programmieren. Ausserdem kann sich keiner so schön ärgern, wenn er beim Kickern ein Tor kassiert hat ☺

Dirk Leusser dafür, dass er es überhaupt erst ermöglicht hat, dass es zu der Kooperation mit dem AK Stalke kam (was dazu führte, dass ich gleich zwei Chefs auf einmal hatte). Ausserdem hat er mir erlaubt, meine Residualdichtetests mit seinen Strukturen durchzuführen. Wie gut, dass wir uns in Florenz kennengelernt haben.

allen Mitgliedern des AK Stalke für eine wunderbar freundschaftliche und entspannte Arbeitsatmosphäre.

Thorsten Müller, stellvertretender Küchenchef der Nordmensa, dafür, dass er das Romanescomodell zur Verfügung gestellt hat.

allen Korrekturlesern, dass sie sich durch die vielen Seiten gequält haben (eventuelle Fehler habe ich also nicht als Einzige übersehen).

meinen Freunden und meiner Familie, vor allem Oli, und ganz besonders **grossen Dank schulde ich meiner Schwester Christina** (Schriftgrösse 14 und fett, so wie du wolltest ☺) punkt. fertig.

Ein Pferd hat vier Beiner,
an jeder Ecken einer,
drei Beiner hätt,
umfallen tät!

*(Fredl Fesl, Liedermacher aus
meiner niederbayrischen Heimat
Niederbayern)*

List of Abbreviations

Å	Ångström (1 Å = 10 ⁻¹⁰ m)
Ala	Alanine
Asp	Aspartic acid
χ	extinction coefficient
CSD	Cambridge Structural Database
Cu	copper
Cys	Cysteine
d	resolution (in Å)
ΔF	$F_{obs} - F_{calc}$
d^f	fractal dimension
Dha	2,3-didehydroalanine
Dhb	(Z)-2,3-didehydrobutyrine
ε	boxlength
e	electron
e_{gross}	gross residual electrons
e_{net}	net residual electrons
Eq.	Equation
<i>et al.</i>	et alii (and others)
\mathbf{F}_{calc}	calculated structure factor with the amplitude F_{calc}
\mathbf{F}_{obs}	observed structure factor with the amplitude F_{obs}
Fig.	Figure
Glu	Glutamic acid
Gly	Glycine
IAM	Independent Atom Model
k	scale factor
K	Kelvin
λ	wavelength
Lan	(2S,6R)-lanthionine
Leu	Leucine

m	meter
M	mol/l
MAD	multi-wavelength anomalous dispersion
MeLan	(2S,3S,6R)-3-methylanthionine
mg	milligram
MIR	multiple isomorphous replacement
μ l	microliter
ml	milliliter
MM	Multipole Model
Mo	molybdenum
PEG 6000	Polyethylene glycol with a mean molecular weight of 6000 g/mol
pH	power of hydrogen
Phe	Phenylalanine
Φ	phase angle
Pro	Proline
QTAIM	Quantum Theory of Atoms in Molecules
ρ	electron density
ρ_0	residual density
RMSD	Root Mean Square Deviation
s	second
SAD	single-wavelength anomalous dispersion
Ser	Serine
SIR	single isomorphous replacement
^t Bu	<i>tert</i> -butyl
Thr	Threonine
TRIS	Tris(hydroxymethyl)-aminomethan
Trp	Tryptophan
Xaa	any amino acid

Contents

Summary	4
Zusammenfassung	6
1 Introduction	8
2 Theoretical Background	10
2.1 X-ray crystallography	10
2.2 Multipole formalism	15
3 Residual Density Analysis	18
3.1 Quantification of the residual density distribution	18
3.1.1 Histograms	19
3.1.2 Fractal dimension	21
3.1.2.1 Mathematical fractals	21
3.1.2.2 Fractals in nature	28
3.1.3 The dimension of fractals	31
3.1.4 The fractal dimension of the residual density	33
3.2 Development of new descriptors	34
3.2.1 Gross residual density and gross residual electrons	34
3.2.2 Net residual density and net residual electrons	35
3.3 Implementation of the new descriptors	36
3.3.1 jnk2RDA	36
3.3.2 Box counting algorithm	38
3.3.3 Line-counting algorithm	39
3.3.4 Optimization of the program jnk2RDA	40
3.3.5 Example output file from jnk2RDA	41
3.4 The prediction of the number of gross residual electrons in the whole unit cell	44
3.5 Results and Discussion	44
3.5.1 Application of new descriptors to simulated data	44

3.5.1.1	Influence of the crystallographic resolution	44
3.5.1.2	Influence of Gaussian noise	46
3.5.1.3	Analysis of the impact of change of model parameters on the residual density	48
3.5.2	Application of new descriptors to experimental data	63
3.5.2.1	Progression of a multipole refinement	63
3.5.2.2	Analysis of the impact of a change of model parameters on the residual density	68
3.5.2.3	Influence of the truncation of the experimental resolution	80
3.5.2.4	Extinction effects	82
3.5.2.5	Grid size effects	85
3.5.2.6	Grid resolution effects	88
3.5.3	Residual density descriptors at the density model and parameter optimum	89
3.5.4	Analysis of the refinement of a disordered and non-centrosymmetric structure	90
3.5.5	Correction for negative intensity observations	93
4	Labyrinthopeptin A2	97
4.1	Introduction	97
4.2	Lantibiotics	98
4.3	Crystallization	102
4.3.1	Basics and principles in crystallization of macromolecules . . .	102
4.3.1.1	Hanging drop vapor diffusion method	103
4.3.1.2	Sitting drop vapor diffusion method	103
4.3.1.3	Cryocrystallography	104
4.3.2	Crystallization conditions for A2	104
4.4	Data processing and structure	105
4.4.1	Nomenclature of the atoms in the residues in A2	108
4.4.2	Ramachandran plot	110
4.4.3	<i>Cis</i> peptide bonds	110
4.4.4	Quaternary D amino acids	116
4.4.5	Comparison of the rings in A2	117
4.4.5.1	Overlay of the rings A and A'	117
4.4.5.2	Overlay of the rings B and B'	118
4.4.5.3	Overlay of the rings AB and A'B'	119
4.4.6	Surface	120

4.4.7	Charge	121
4.4.8	Hydrogen bonds	122
4.4.9	Crystal packing	123
5	List of developed programs	125
	Bibliography	127
	Appendix	135
A.1	The net residual electrons in the unit cell	135
A.2	Gross residual electrons in the whole unit cell derived from the Fourier coefficients of the residual density	136
A.3	The gross residual electrons in the unit cell, in a part of the unit cell and standard quality measures	139
A.4	Prediction of $d^f(0)$ from experimental resolution and grid resolution .	140
A.4.1	Derivation	141
A.4.2	Discussion	142
	List of Publications	145
	Poster presentations	148
	Curriculum vitae	149

Summary

The present work is divided into two main parts. The first part deals with the analysis and validation of the residual density. For the quantification and analysis the concept of the fractal dimension is employed. For this purpose several new descriptors were developed and implemented into a program called “jnk2RDA”. For testing the influence of the experimental resolution on the descriptors ideal simulated data were created to exclude the influence of any unknown sources of error. It could be shown that the fractal dimension of the residual density is strongly dependent on the resolution of the data. For simulating a more “realistic” case Gaussian noise was added to the ideal data. It was seen that the fractal dimension of the zero residual density is relatively independent on the amount of noise whereas the flatness (maximum and minimum residual density values) decreases with increasing noise.

Several parameters describing the electron density were varied from their refined values for simulated and for experimental data to investigate their influence on the residual density. It could be shown that each parameter has its individual influence on the distribution of the residual density and on its fractal dimension distribution and that the shape of the fractal dimension distribution can help to identify which parameter could be set to a more appropriate value. It was seen that the manipulation of parameters for experimental data does not necessarily show the expected results as the sources of error cannot always be identified and controlled.

The progression of a multipole refinement starting from the Independent Atom Model to the final Multipole Model was monitored with the residual density descriptors and it could be shown that the parameters which contribute most to the improved description of the electron density are the monopole and multipole population parameters. An investigation of the influence of extinction correction on the residual density showed that the refinement of an even relatively small extinction parameter can improve the residual density remarkably. This improvement can be very large compared to any model improvement even including what can be achieved by refining the multipole population parameters.

It was shown that the size and resolution of the residual density grid has an enormous impact on the residual density descriptors. The optimum grid has a

resolution that is in a certain relation to the experimental resolution and which is proportional to the cell axes. For properly chosen values the computing effort and memory demand are minimized and the obtained information is maximized.

The residual density descriptors were applied to verify the correct refinement on data of a disordered structure with multipole methods. It was shown that the residual density did not only improve in the molecular region when the disorder is taken into account but also for the whole unit cell.

A new method was developed which allows for the correction of negative intensity observations that can occur for several reasons in an X-ray crystallography measurement. This new algorithm is a very effective way to correct the data and to make it possible that also negative observed intensities are included in the refinement, which is especially important for high resolution data. No such treatment is existing yet for small molecule crystallography.

The second part of this work presents a new lantibiotic (called labyrinthopeptin A2) consisting of 18 amino acids, which shows new structural features. This cyclic peptide contains the amino acid lanthionine, two unusual *cis* peptide bonds and new intramolecular links. The C_α atoms of two alanine residues are bonded to the C_β atoms of two neighbored alanine residues. This leads to quaternary substituted C_α atoms which is an uncommon structural motif. Despite the small rings in A2 consisting of only four amino acids the dihedral angles lie all in allowed regions in the Ramachandran plot. The peptide can formally be divided into two segments each containing two rings, the main chains of which are structurally very similar, which was shown by an overlay of the corresponding atoms.

Zusammenfassung

Die vorliegende Arbeit ist in zwei Teile untergliedert. Der erste Teil beschäftigt sich mit der Analyse und Überprüfung der Residualdichte. Für die Quantifizierung und Analyse wurde das Konzept der fraktalen Dimension verwendet. Zu diesem Zweck wurden verschiedene neue Deskriptoren entwickelt und in ein Programm namens “jnk2RDA” implementiert. Um den Einfluss der experimentellen Auflösung auf die Deskriptoren zu untersuchen wurden ideale Daten erzeugt um unbekannte Fehlerquellen auszuschliessen. Es konnte gezeigt werden, dass die fraktale Dimension der Residualdichte stark von der Auflösung der Daten abhängt. Um realistischere Bedingungen zu simulieren wurden die idealen Daten Gauss-verrauscht. Es zeigte sich, dass die fraktale Dimension der Nullresidualdichte relativ unabhängig von der Menge des Rauschens ist, wogegen die Breite der Verteilung (maximale und minimale Restdichtewerte) mit ansteigendem Rauschen abnimmt.

Verschiedene Parameter, die die Elektronendichte beschreiben, wurden ausgehend von ihren verfeinerten Werten für ideale und für experimentelle Daten variiert. Es konnte gezeigt werden, dass jeder Parameter einen individuellen Einfluss auf die Verteilung der Residualdichte und auf die Verteilung von deren fraktaler Dimension hat. Ausserdem kann die Form der Kurve der fraktalen Dimension bei der Identifizierung von Parametern helfen, die unter Umständen noch weiter angepasst werden können. Es zeigte sich, dass die Änderungen der Parameter für experimentelle Daten nicht immer die erwarteten (den idealen Daten entsprechenden) Ergebnisse zeigen, da in experimentellen Daten meist Fehler enthalten sind, die nicht immer identifiziert und kontrolliert werden können.

Der Verlauf einer Multipolverfeinerung, ausgehend vom IAM Modell bis hin zum finalen Multipolmodell, wurde mit den Residualdichtedeskriptoren beobachtet, und es zeigte sich, dass die Parameter, die den grössten Beitrag zur Verbesserung der Beschreibung der Elektronendichte liefern, die Mono- und Multipolpopulationsparameter sind. Eine Untersuchung des Einflusses der Extinktionskorrektur auf die Residualdichte zeigte, dass sogar die Verfeinerung von relativ kleinen Extinktionsparametern die Residualdichte enorm verbessern (reduzieren) kann. Diese Verbesserung

kann sogar deutlich grösser sein als diejenige, die durch Einführung der Multipolparameter erzielt wird.

Es konnte gezeigt werden, dass die Auflösung des Residualdichtegrids die Residualdichtedeskriptoren beeinflusst. Die optimale Gridauflösung hängt von der experimentellen Auflösung ab und ist proportional zu den Zellkanten. Für optimal gewählte Werte kann bei gleichbleibendem Informationsgehalt der rechnerische Aufwand minimiert werden.

Eine Anwendung der Residualdichtedeskriptoren war die Bestätigung der Multipolverfeinerung einer fehlgeordneten Struktur. Es konnte gezeigt werden, dass sich unter Berücksichtigung der zweiten Molekülposition nicht nur die Residualdichte in der Nähe der Atome verbessert hat sondern in der gesamten Einheitszelle.

Ausserdem wurde eine neue Methode entwickelt, mit der negativ beobachtete Intensitäten korrigiert werden können, die bei einer Röntgenstrukturanalyse jederzeit auftreten können. Dieser neue Algorithmus kann die Daten auf eine sehr effektive Art korrigieren und ermöglicht die Verwendung aller (auch negativ beobachteter) Daten bei der Verfeinerung, was vor allem bei hochaufgelösten Messungen wichtig ist. Bisher gibt es solche Korrekturen nur für niedrig-aufgelöste Proteindaten aber nicht in der Kleinmolekülkristallographie.

Im zweiten Teil der Arbeit wird ein neues Lantibiotikum (Labyrinthopeptin A2) vorgestellt, das aus 18 Aminosäuren aufgebaut ist und neue strukturelle Eigenschaften besitzt. Das zyklische Peptid enthält die Aminosäure Lanthionin, zwei ungewöhnliche *cis*-Peptidbindungen und neuartige intramolekulare Verknüpfungen. Dabei sind die C_{α} -Atome von zwei Alaninresten an die C_{β} -Atome von zwei benachbarten Alaninresten gebunden. Das führt zu quartär substituierten C_{α} -Atomen, was in Peptiden und Proteinen ein seltenes Strukturmotiv ist. Trotz der kleinen Ringe in A2, die teilweise nur aus vier Aminosäuren bestehen, liegen die Diederwinkel alle in erlaubten Regionen im Ramachandran-Plot. Das Peptid kann formal in zwei Segmente aufgeteilt werden, die je aus zwei Ringen aufgebaut sind, deren Hauptketten strukturell sehr ähnlich sind, was durch eine Überlagerung der entsprechenden Atome gezeigt werden konnte.

1 Introduction

X-ray crystallography serves as an important tool to investigate the molecular structure of many chemical compounds. Knowing the exact structure is essential to understand the mechanism of reactions. Also, atomic radii, bond lengths and the bonding situation in molecules are determined with the help of X-ray crystallography. Getting a detailed knowledge of the structural situation of a molecule is especially important for macromolecules when information about activity or the mechanism for recognition and binding of substrates is investigated. This helps to show how pharmaceuticals interact with their targets and how the structure can be modified to improve this interaction.

The present work is divided into two main parts. The first part deals with the analysis and validation of high resolution X-ray datasets. In the last years the methods for exact measurements to very high resolution have improved remarkably, while the utilities for characterizing the data have not been adapted to the new requirement. The quality measures are mostly the same as for conventional refinements, where the challenge is the correct assignment of the atom type, but not the determination of the details like the second derivative of the electron density. It is clear that more exact measurements are more and more sensitive towards statistical and systematic errors, which might originate from the measurement, the data processing or modeling or maybe even from too idealized assumptions made by theory. Additionally, some standard procedures like the omission of negative intensity observations have to be reconsidered, as any systematic exclusion of data increases the systematic error. Thus, it is important to have analysis tools allowing for a global as well as a local analysis of the deviations of the model from the data, which may not only be used for the evaluation of the final model, but which also facilitate the decisions on how to proceed at certain stages of the refinement.

During this work a program called jnk2RDA was developed that allows the evaluation of the residual density distribution. Some general background on X-ray crystallography (section [2.1](#)) and on Multipole Methods (section [2.2](#)) is given. As the

program uses the concept of the so-called “fractal dimension” the topic “fractals” will be briefly introduced in section 3.1.2 and it will be shown how their dimension is determined. In section 3.3.1 the program jnk2RDA is briefly introduced and it is described which file formats it can process. Some new descriptors of the residual density were developed which are described in section 3.2. Different algorithms that can be used to evaluate the fractal dimension distribution of the residual density are introduced (see section 3.3.2) one of which was developed during this work. It is shown that this new algorithm is more efficient than the others. In section 3.3.4 the optimization of the program jnk2RDA is described in detail and an example output file is shown (section 3.3.5). The program is tested on simulated ideal data (section 3.5.1) and on experimental data (section 3.5.2) of the same compound. The dependence of the program on the experimental resolution and the grid resolution is investigated. A chapter on the application of the program on data of a disordered structure which was refined with the Multipole Model follows in section 3.5.4. The influence of the incorporation of negative intensities into the refinement was tested for which purpose a program was written that corrects the set of observed reflections for negative intensities (section 3.5.5).

The second part of the work deals with the crystallization and refinement of the crystal structure of a lantibiotic called labyrinthopeptin A2 that shows new structural features. In sections 4.1 and 4.2 antibiotics in general and the biosynthetic pathway of lantibiotics are briefly introduced. The crystallization procedure of A2 with protein crystallization methods is described (section 4.3.2) and the structure solution and refinement that could be performed with small molecule methods (direct methods) due to the relatively high resolution of 1 Å. A2 is a cyclic peptide consisting of five rings. It has two unusual *cis* peptide bonds and two quaternary C_α atoms that are bridged to the C_β atoms of neighboring alanine residues. The peptide shows slight antibiotic, antiviral and more pronounced analgetic activity.

Apart from the main program jnk2RDA more software was developed to allow for the manipulation of the data files. For example `factor_int_sig` adds Gaussian noise to a list of observed intensities and `histomatch_fco` corrects high resolution data from negative intensity observations. A more detailed description and an overview of these programs is given in section 5.

2 Theoretical Background

2.1 X-ray crystallography ^[1]

X-rays are electromagnetic waves with wavelengths of about $\lambda = 0.5 - 3.0 \text{ \AA}$. They are produced when electrons of high energy, which were accelerated from a cathode towards an anode of a certain material like Mo or Cu, suddenly hit the anode. This leads to the emission of X-rays of characteristic wavelengths. In a crystal the atoms are arranged regularly in repeating units. The smallest unit which builds up the crystal just by translation is called the unit cell. The monochromated X-rays are directed towards the crystal where they are scattered by the electrons in the molecules. As the X-rays are considered as waves they can be described with an amplitude F and a phase Φ .

$$\mathbf{F} = |\mathbf{F}| e^{i\Phi} = F e^{i\Phi} \quad (2.1)$$

The amplitude of the vector \mathbf{F} is its modulus $|\mathbf{F}|$ and is denoted F . The amplitude contains information on the amount of electrons present in the unit cell, whereas the phase contains information on their relative position to each other. For convenience reasons the scattered waves are considered as reflections at so-called lattice planes. Only for certain angles θ between the X-ray source and the lattice planes reflections can be observed. This relation is given by the Bragg equation ^[2]

$$n\lambda = 2d \sin \theta \quad (2.2)$$

with $n =$ an integer number, $\lambda =$ wavelength, $d =$ distance between the lattice planes (resolution), $\theta =$ scattering angle. The detected reflections are described with their indices h , k and l that denote their orientation relative to the unit cell, and their intensity I . According to the kinematical theory of scattering ^[3] the intensity I of

the reflections is proportional to the square of the amplitude F

$$I \propto F^2 \quad (2.3)$$

Unfortunately, the phase information is lost during the measurement. This is called the phase problem in crystallography. Several methods that have been developed to solve this phase problem will be described later.

As the beam is scattered by all atoms in the crystal each atom contributes to each reflection. Thus, the structure factor \mathbf{F}_{hkl} can be expressed as the summation over all atomic structure factors \mathbf{F}_i . The individual atomic structure factors \mathbf{F}_i are composed of the atomic scattering factor f_i and the phase shift Φ_i caused by this atom. Fig. 2.1 shows a representation of an atomic scattering factor in the Argand diagram. In the case of atomic structure factors \mathbf{F}_i the amplitude is the atomic scattering factor f_i . For a scattering angle $\theta = 0^\circ$ the atomic scattering factor f_i is normalized to the number of the electrons of the atom and it decreases with increasing resolution.

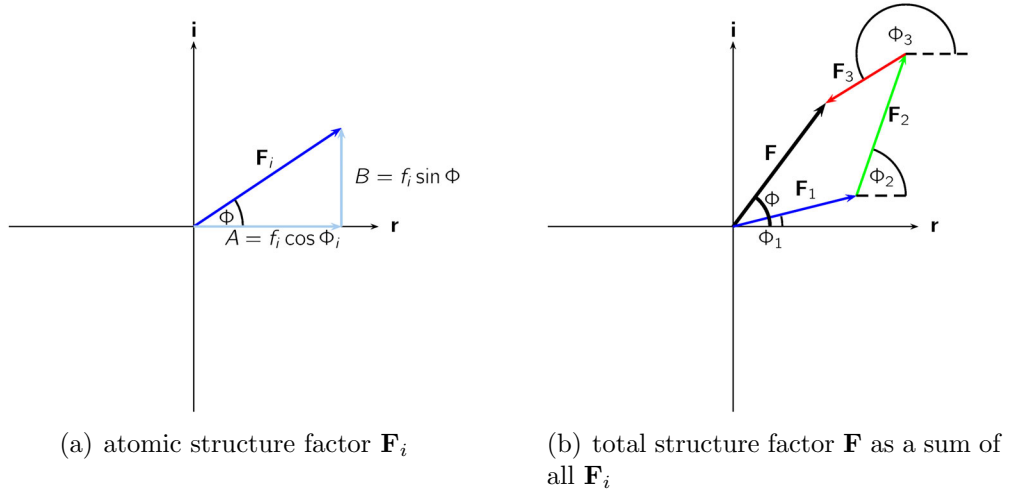


Figure 2.1: Projection of the structure factor \mathbf{F} in the Argand diagram.

$$\mathbf{F}_{hkl} = \sum_i f_i \{ \cos[2\pi(hx_i + ky_i + lz_i)] + i \sin[2\pi(hx_i + ky_i + lz_i)] \} \quad (2.4)$$

h , k and l are the Miller indices of the lattice planes and x_i , y_i and z_i are the coordinates of atom i . The structure factor can also be calculated from the contribution

of the electron density ρ in small volume elements. The Fourier transformation is

$$\mathbf{F}_{hkl} = |\mathbf{F}_{hkl}| e^{i\Phi_{hkl}} = \int_V \rho(xyz) e^{2\pi i(hx+ky+lz)} dV \quad (2.5)$$

When the Fourier backtransformation is applied the electron density can be calculated from the summation over all individual structure factors in the volume V with

$$\begin{aligned} \rho(xyz) &= \frac{1}{V} \sum_{hkl} \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \\ &= \frac{1}{V} \sum_{hkl} F_{hkl} e^{i\Phi_{hkl}} e^{-2\pi i(hx+ky+lz)} \end{aligned} \quad (2.6)$$

As already mentioned the amplitudes F_{hkl} can be calculated from the measured intensities of the reflections, but the information on the phase Φ is lost.

“Direct methods” are one of several possibilities to reconstruct the phases. They can be applied mainly for small molecules with up to about 1000 atoms and when data up to atomic resolution are ($d \leq 1.2 \text{ \AA}$) are available. Direct methods are based on the relation between intensities and phases and the assumption that the electron density is always positive and that its maxima are distributed evenly in the crystal. A fundamental contribution to direct methods had the Sayre equation ^[4] from which the triplet phase relation can be derived ^[5,6]. It is based on the relation between phases of three independent but strong reflections

$$\Phi_{h+h',k+k',l+l'} \approx \Phi_{h,k,l} + \Phi_{h',k',l'} \quad (2.7)$$

Direct methods were developed in more detail by Karle and Hauptman ^[7-9]. From the initial phases an initial model can be calculated.

Other methods for the determination of phases for larger molecules like proteins are the isomorphous replacement methods like SIR or MIR, which allow to get the phases from a heavy atom derivative. Anomalous dispersion methods are based on the anomalous signal that heavy atoms show for certain wavelengths (SAD and MAD). If a structurally similar compound is available Molecular Replacement can be the method of choice.

The structure factors derived from the observed intensities I are denoted \mathbf{F}_{obs} and the calculated structure factors \mathbf{F}_{calc} . They are scaled to each other by the scale

factor k

$$k = \frac{\sum_{hkl} F_{obs}^2}{\sum_{hkl} F_{calc}^2} \quad (2.8)$$

This scale factor is always necessary when the F_{calc} are set into relation to the F_{obs} but it will not specifically be mentioned in the following equations.

With the initial model and phases an improved model can be calculated by a least-squares refinement of the model parameters, from which new and improved amplitudes and phases are obtained. This process is iteratively repeated. The model is calculated such that the squared difference in the weighted squared amplitudes minimizes if the refinement is performed against F^2 .

$$\sum_{hkl} w(F_{obs}^2 - F_{calc}^2)^2 = Min. \quad (2.9)$$

The factor w weights the intensities according to their accuracy and reliability, *i.e.* the uncertainties in their measurement are accounted for. In the simplest case the weights w are

$$w = \frac{1}{\sigma^2(F_{obs}^2)} \quad (2.10)$$

In the refinement program SHELXL^[10,11] this weighting scheme is extended to minimize the differences in the squares of the errors for reflections in different intensity ranges.

$$w = \frac{1}{\sigma^2(F_{obs}^2) + (aP)^2 + bP} \quad (2.11)$$

with

$$P = \frac{2F_{calc}^2 + Max(F_{obs}^2, 0)}{3} \quad (2.12)$$

The values for a and b are proposed by SHELXL during the refinement.

For the minimization of the sum of the squared differences of the squared structure factors different parameters for each atom have to be refined. For a normal small molecule Independent Atom Model (IAM) these are nine parameters per atom (three coordinates x , y , z and six anisotropic displacement parameters U_{ij}).

For a comparison how well the model agrees with the observed data R -factors are calculated. The conventional R_1 -factor is based on a comparison of the amplitudes F

$$R_1 = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|} \quad (2.13)$$

If the R_1 is weighted the wR_1 is obtained:

$$wR_1 = \sqrt{\frac{\sum_{hkl} w ||F_{obs}| - |F_{calc}||^2}{\sum_{hkl} w |F_{obs}|^2}} \quad (2.14)$$

As the refinement is usually performed against the squared amplitudes F^2 an R_2 -factor can be calculated as

$$R_2 = \sqrt{\frac{\sum_{hkl} (F_{obs}^2 - F_{calc}^2)^2}{\sum_{hkl} (F_{obs}^2)^2}} \quad (2.15)$$

Normally, the weights are applied in the refinement, thus, the weighted wR_2 is given as

$$wR_2 = \sqrt{\frac{\sum_{hkl} w (F_{obs}^2 - F_{calc}^2)^2}{\sum_{hkl} w (F_{obs}^2)^2}} \quad (2.16)$$

An additional quality criterion is the goodness of fit $Goof$.

$$Goof = S = \sqrt{\frac{\sum_{hkl} w (F_{obs}^2 - F_{calc}^2)^2}{(n - p)}} \quad (2.17)$$

with n = number of reflections and p = total number of parameters refined. The $Goof$ not only considers the weighted error in the squared difference $F_{obs}^2 - F_{calc}^2$ but also the degree of overdetermination of the parameters with the difference $(n - p)$. If the structure was refined completely and correctly and the weighting scheme was applied properly the $Goof$ should give a value close to 1.

For unproblematic small molecule structures the R_1 -factor can reach values smaller than 0.05 and the wR_2 values less than 0.15.

Another R -factor that is used in this work is the R_{int} . It is calculated from the summation over all reflections which are averaged over at least one symmetry equiv-

alent.

$$R_{int} = \frac{\sum |F_{obs}^2 - \langle F_{obs}^2 \rangle|}{\sum F_{obs}^2} \quad (2.18)$$

The R_{int} describes the deviation of a reflection from its symmetry equivalents.

In X-ray crystallography there are several approaches to describe the electron density in a crystal. One of them is the conventional Independent Atom Model (IAM). It describes a spherical electron density around the atoms and does not account for charge transfer and bonding between atoms. Thus, quite high residual electron densities remain after an IAM refinement. The higher the resolution of the experiment the more distinct are these residuals.

2.2 Multipole formalism

The multipole approach accounts for the distribution of the electrons in bonds and in lone pairs. It also allows for charge transfer between atoms. It uses an aspherical model for the description of the electron density based on the nucleus centered finite multipole expansion which was first developed by Stewart ^[12-15] and modified by Hansen and Coppens ^[16]. This formalism is implemented into the program package XD ^[17], from which mainly the least-squares program XDLSM, the Fourier transformation program XDFOUR and the graphics program XDGRAPH were used in this work.

In the multipole model the total molecular electron density is decomposed into pseudoatomic contributions. Each pseudoatom is further decomposed into the following components

$$\rho_{atomic}(\mathbf{r}) = \rho_c + P_\nu \kappa^3 \rho_v(\kappa r) + \rho_d(\kappa' \mathbf{r}) \quad (2.19)$$

where ρ_c is the (frozen) core density, ρ_v is the spherical valence density and ρ_d is the deformation density. P_ν are the monopole population parameters and the κ and κ' values account for the radial expansion and contraction of the spherical valence and deformation density, respectively.

The deformation density can be described as

$$\rho_d(\kappa' \mathbf{r}) = \sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa' r) \sum_{m=-l}^l P_{lm} y_{lm}\left(\frac{\mathbf{r}}{r}\right) \quad (2.20)$$

where the R_l describe the radial distribution of the electron density

$$R_l(r) = \frac{\zeta^{n_l+3}}{(n_l+2)!} r^{n_l} e^{-\zeta r} \quad (2.21)$$

and the y_{lm} are normalized real spherical harmonic functions which describe the angular dependence of the density. P_{lm} are the multipole population parameters. The coefficients n_l are positive integer values with $n_l \geq l$ according to Poisson's equation ^[18] and the ζ are energy-optimized single-Slater orbital exponents. ^[19,20]

Additional to the three positional and the six harmonic vibration parameters that are to be refined in the Independent Atom Model, in the Multipole Model the parameters for the monopole and multipole populations P_ν and P_{lm} , respectively, and the radial expansion/contraction parameters κ and κ' can be refined, which allows a much more accurate description of the measured electron density.

Fig. 2.2 shows a graphic representation of the multipolar functions y_{lm} according to Eq. 2.20. The radial part $R_l(r)$, the expansion/contraction parameter κ' and the multipole population parameters P_{lm} were set to 1 for the calculation. For the monopolar function $l = 0$, for the dipolar functions $l = 1$, for the quadrupolar functions $l = 2$, for the octupolar functions $l = 3$ and for the hexadecapolar functions $l = 4$. m is going from $-l$ to $+l$ which leads to one monopolar, three dipolar, five quadrupolar, seven octupolar and nine hexadecapolar functions y_{lm} .

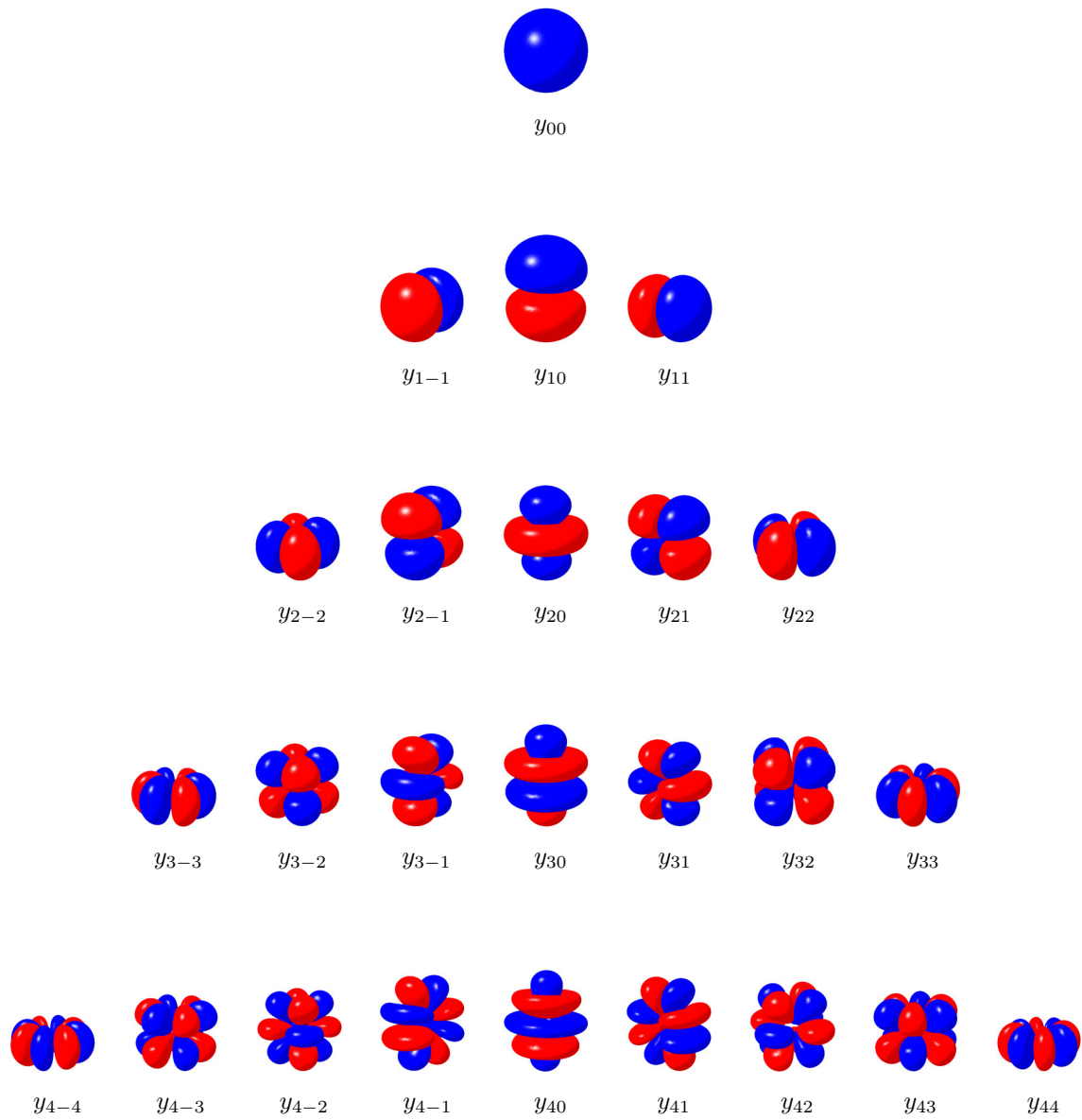


Figure 2.2: Multipolar functions y_{lm} describing the valence deformation density; monopolar function: y_{00} , dipolar functions: y_{1m} , quadrupolar functions: y_{2m} , octupolar functions: y_{3m} , hexadecapolar functions: y_{4m} ; Figures were generated with Molliso ^[21].

3 Residual Density Analysis

3.1 Quantification of the residual density distribution

At the end of a least-squares refinement on X-ray data there must always be a check or better several different independent checks on the results. Possible errors must be uncovered and the correctness of the model must be assured. For proteins and peptides, where the resolution is relatively low and the depth of detail in the structure is not very high, the quality measures differ from the methods for evaluation of the data for moderate (high) resolution measurements of small molecules.

The usual quality measures like the R -factors (Eqs. 2.13, 2.16) are global measures of quality. They describe, how the model as a whole fits the experimental data. The agreement between theory and experiment increases, when *e.g.* a wrong atom type is replaced by the correct type. This usually also changes the vibration parameters. In high resolution experiments, however, the changes are much more subtle. While the replacement of an atom is a discontinuous act, which also changes the total scattering power emerging from that particular point in the unit cell, in high resolution experiments the question is rather of the type if there are 4.4 or 4.6 electrons to be assigned to that scatterer, for example. This number can change in a continuous way. Moreover, parameter correlations *e.g.* between the mentioned monopole population and thermal vibration, make it difficult to determine this number. It is, however, easy to imagine that those small model parameter changes often affect the total R -factors only moderately. On the other hand, it is exactly these parameters, which after a converged refinement themselves determine the physical and chemical interpretation of the electron density distribution (*e.g.* within the framework of QTAIM). For this reason it would be helpful to allow the quality measures to become local also, such that the minor changes in the overall R -factors can be magnified and such that a change in the residual density distribution due to a model parameter change can be visualized and quantified.

All of the following new developed residual density descriptors can be evaluated for the unit cell as a whole as well as for any subvolume of the unit cell. Moreover,

the ideal values (the limiting values for the ideal case of $F_{obs} = F_{calc}$ for the whole experimental resolution) for each subvolume as well as for the whole unit cell can be given easily. In this way, the total error can be decomposed in errors emerging from the different parts of the unit cell. In this sense, the descriptors are global and local. In the following paragraphs these residual density descriptors are introduced.

The residual density is the difference density for observed data \mathbf{F}_{obs} and calculated data \mathbf{F}_{calc} when the applied phase Φ is the calculated phase Φ_{calc} for both sets of structure factors. The residual density ρ_0 is defined as

$$\rho_0(xyz) = \frac{1}{V} \sum_{hkl} (F_{obs} - F_{calc}) e^{i\Phi_{calc}} e^{-2\pi i(hx+ky+lz)} \quad (3.1)$$

Two important properties that are required from the residual density for a successful refinement are flatness and featurelessness. The flatness describes the range of the residual density in terms of its maximum and minimum values (highest peak and deepest hole). It can be easily calculated by a peak search analysis of the residual density and should of course be as small (flat) as possible. But the determination of the featurelessness is so far based on highly subjective criteria. It is determined by a visual inspection of certain planes in the unit cell which contain the heaviest atoms of the structure and when the distribution of the residual density seems to be statistical it is called featureless. Unfortunately, features in the vicinity of lighter atoms or in regions where no atoms are expected to be can be overlooked and the residual density is misleadingly termed featureless. Moreover, no number is assigned to the property featureless. Thus, it is necessary to find a possibility to characterize the residual density for the whole unit cell and quantify it for comparison of one model with the other or different structures with each other.

3.1.1 Histograms

One possibility to inspect the residual density for the whole unit cell is to plot a histogram for the distribution of the residual density values. Fig. 3.1 shows such a histogram for a residual density distribution of $S(N^tBu)_3$ [22].

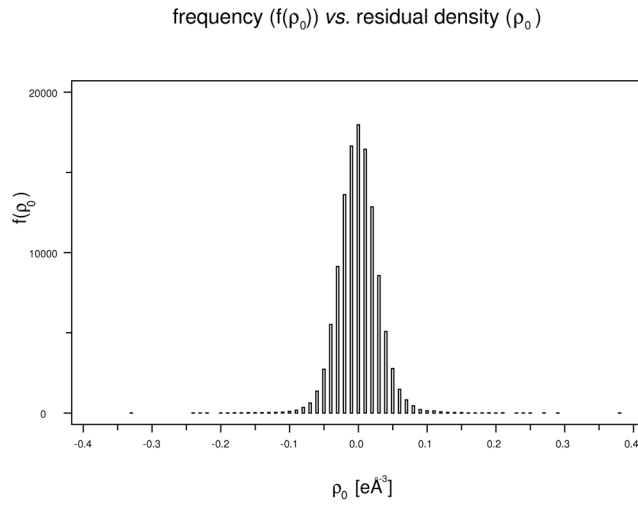


Figure 3.1: Histogram representation of the residual density distribution (see section 5).

For the distribution of the residual density a Gauss function is assumed to be a valid approximation if no systematic error is present [23]. The histogram in Fig. 3.1 does look Gaussian-like as it has one maximum, it is symmetrically distributed and the shape changes from convex to concave and back (bell-shaped). The amount of the residuals in the periphery, however, is not apparent due to the high frequency of residuals close to zero.

Thus, a logarithmic scale of the frequency of the residuals might reveal the details for the high and low residual density values. Fig. 3.2 shows the same histogram as above on a logarithmic scale for the frequency.

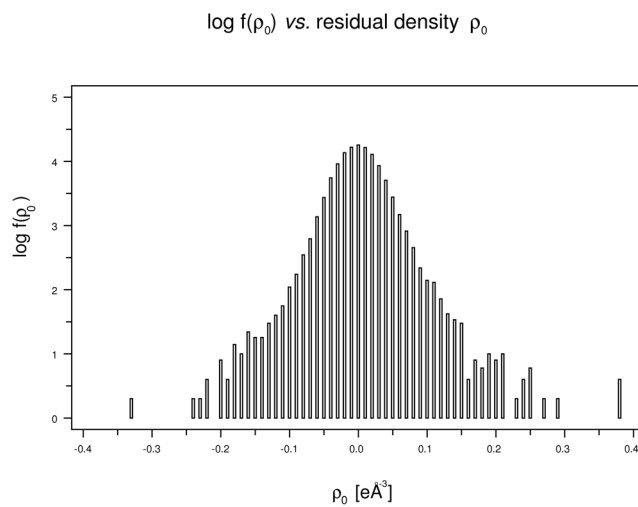


Figure 3.2: Logarithmic histogram representation of the residual density distribution (see section 5).

By taking the logarithm of a Gauss function the resulting graph should be a parabola. In contrast to Fig. 3.1 it can now be seen in Fig. 3.2 that the underlying distribution is far from being a Gaussian due to features in the periphery. There are systematic errors present in the residual density which can be seen in the deviation of the graph from a parabolic shape especially in the region of the very high (and low) residual density values. A problem that remains is the upper boundary of the histogram. The maximum value of about 4.3 in this case has no special meaning and might change for a different structure or even for the same structure if the residual density grid size is varied. Comparability of different structures or models is not given, thus, another descriptor for the residual density and its featurelessness has to be found.

This problem can be solved when the residual density (at different isovalues) is regarded as a fractal and its dimension is calculated.

3.1.2 Fractal dimension

The term fractal was coined by Mandelbrot (from the latin word frangere, fractus: to break, broken) ^[24,25] who found that those irregularly shaped objects have merited interest although they cannot be described with the usual measures because their geometry cannot be described with the usual Euclidean dimensions.

3.1.2.1 Mathematical fractals

Fractals are often characterized by a high degree of self-similarity, which means that each detail from the original object can also be seen at any degree of magnification ^[26]. They have infinite length and are constructible by iteration.

Cantor dust One of the most famous fractals is the Cantor dust ^[27-31]. It is generated by starting from a line with unit length $[0,1]$ and iteratively taking out the middle third of each resulting line. This can be done infinitely often and the object decreases from a line to an infinite set of points (Fig. 3.3). Thus, the dimension of the object is not that of a line anymore, but is more than that of a point as it consists of an infinite number of points.

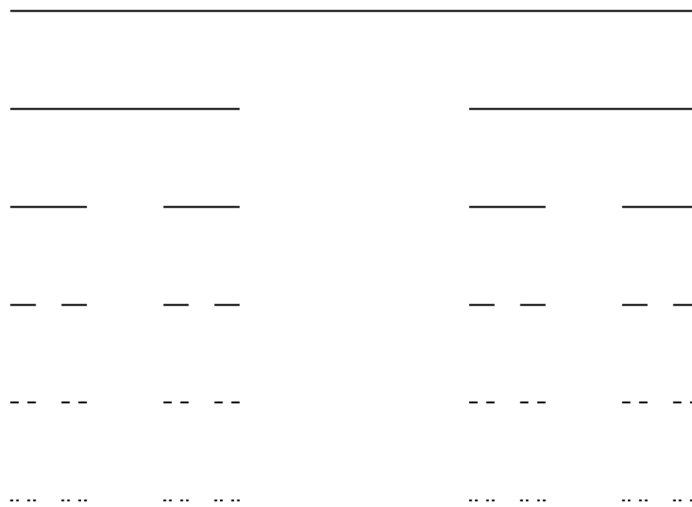


Figure 3.3: Iterative procedure of generating the Cantor dust with initiator (top) and generator (second line) applied iteratively five times.

Sierpinski triangle Another very famous fractal is the Sierpinski gasket or Sierpinski triangle ^[32,33]. It can be generated by starting from a triangle and subsequently removing the middle equilateral triangle of the newly generated triangle (Fig. 3.4). The perimeter of the resulting white area thus increases to infinity.

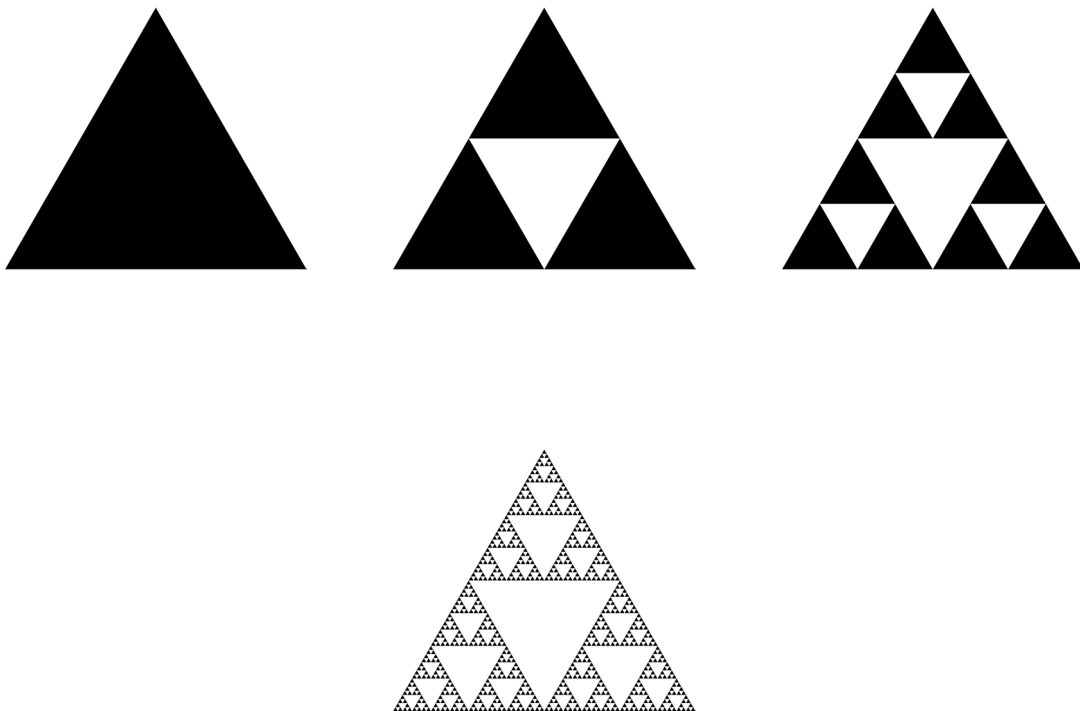


Figure 3.4: Iterative procedure of generating the Sierpinski triangle; the initiator is the black triangle on the left hand side; the generator is applied iteratively six times; steps 3–5 are not shown.

Sierpinski carpet Very similar to the Sierpinski gasket is the Sierpinski carpet ^[32,33]. It is constructed by starting from a square from which the middle square is removed. This is repeated for the resulting squares *ad infinitum* (Fig. 3.5). Here again, the area of the white parts has an infinitely large circumference.

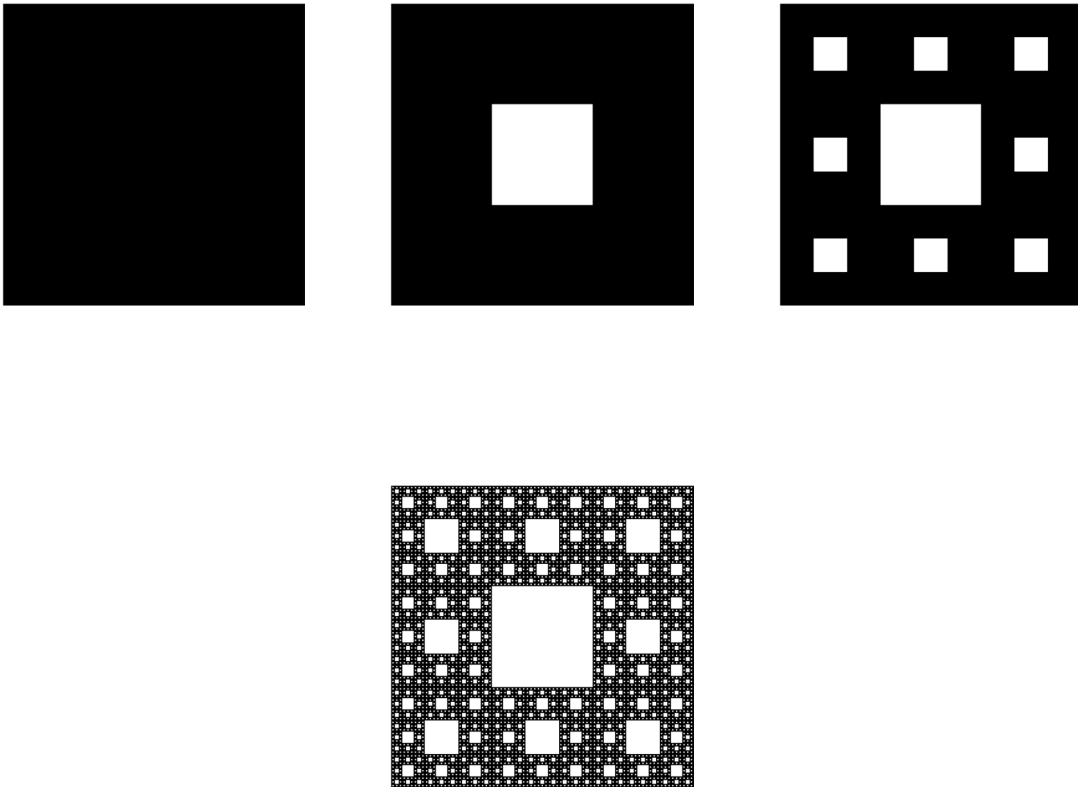


Figure 3.5: Iterative procedure of generating the Sierpinski carpet; the initiator is the black cube on the left hand side; the generator is applied iteratively five times; steps 3 and 4 are not shown.

Koch curve Another famous fractal is the Koch curve from which the Koch snowflake can be constructed ^[34,35]. The Koch curve is generated by subsequently replacing the middle third of a line with the corresponding other two sides that form an equilateral triangle (Fig. 3.6). Three Koch curves put together form the Koch snowflake (Fig. 3.7).

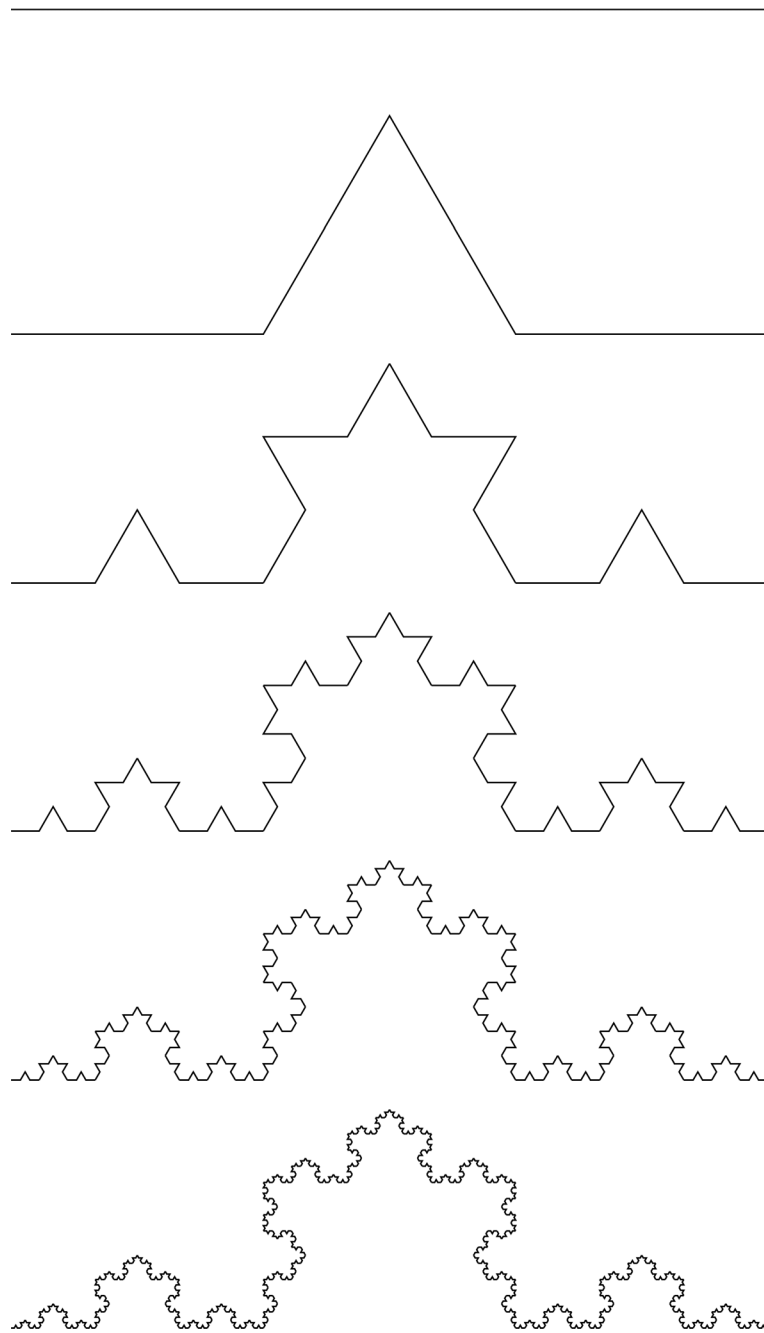


Figure 3.6: Iterative procedure of generating the Koch curve; the initiator is the topmost line; the generator is applied iteratively five times.

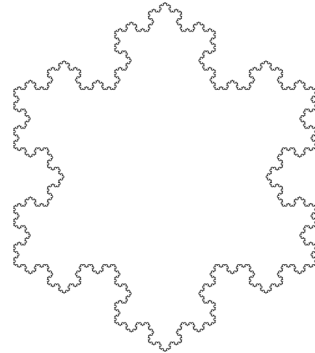


Figure 3.7: Koch snowflake generated from three Koch curves.

Peano curve The Peano curve is a continuous curve that completely fills a 2-dimensional plane ^[36]. It is built up by dividing a line into three segments and putting a square up and below the middle segment.

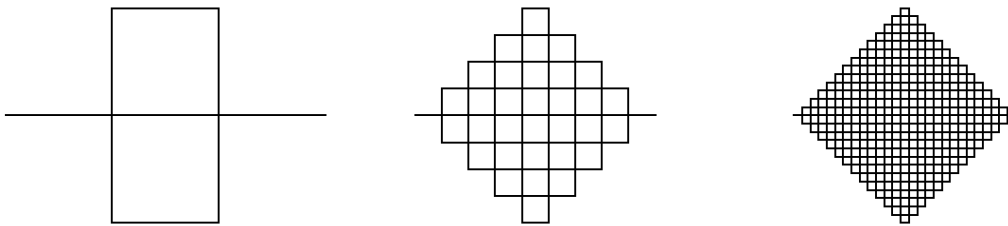


Figure 3.8: Peano curve. ^[37]

Mandelbrot set A really beautiful example of artificial fractals is the Mandelbrot set ^[24,25]. It is constructed in the complex plane by an iteration according to

$$z_{n+1} = z_n^2 + c \quad (3.2)$$

where c is a complex number and the starting value z_0 is 0. The iteration is processed for the complex plane *e.g.* from $c = (-2 - 2i)$ to $c = (2 + 2i)$. If after a certain number of iterations the resulting value for z converges to a constant value, *i.e.* the sequence is bounded, this point belongs to the Mandelbrot set, otherwise, if the sequence diverges, it does not. In Fig. 3.9 the values of the Mandelbrot set are drawn in black and the color coding of the other values describes the number of iterations that were necessary to exceed a threshold value that was set as boundary.

The magnification of the last image in Fig. 3.9 relative to the first one is about $10^{10} : 1$.

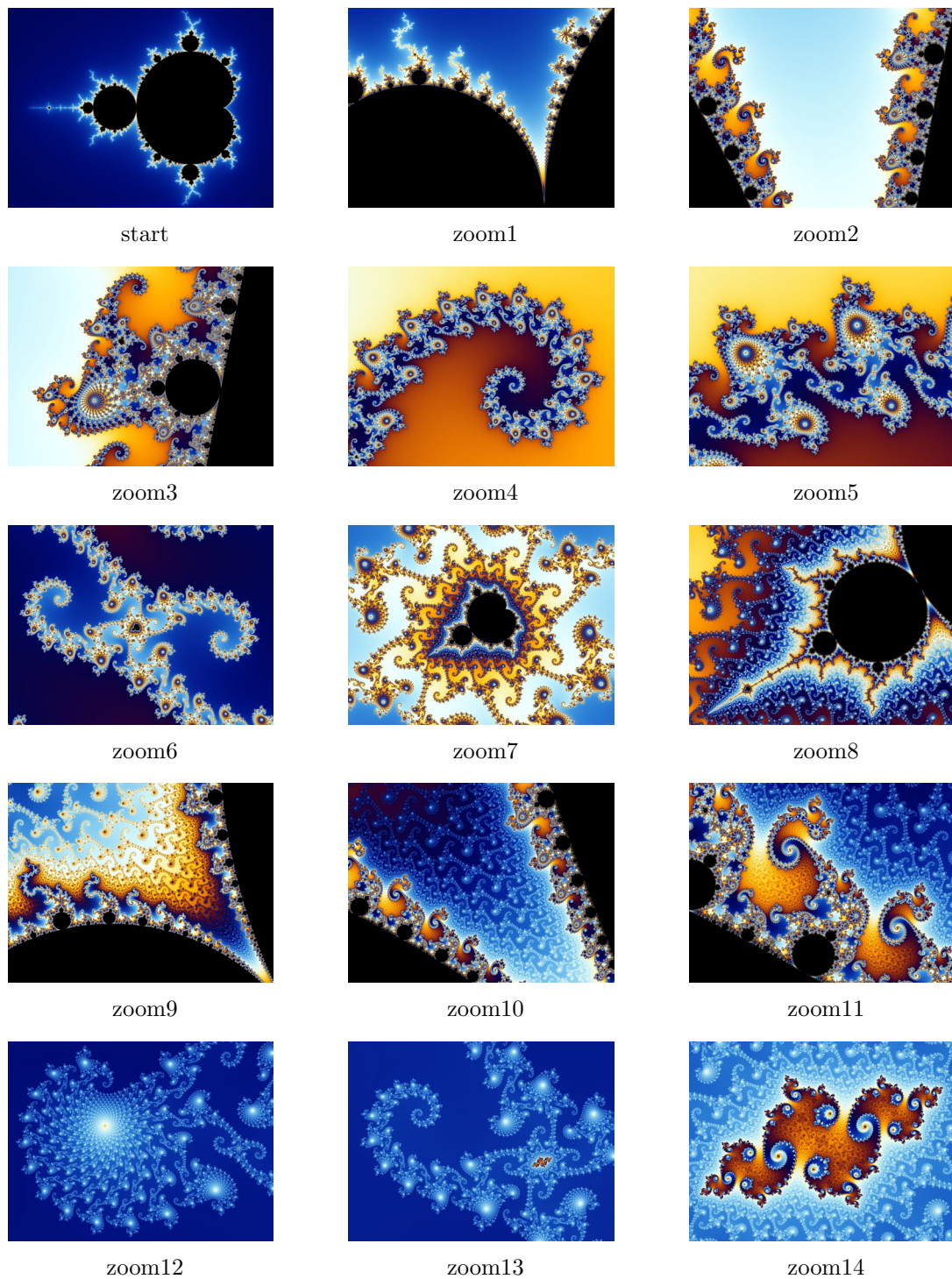


Figure 3.9: Mandelbrot set with 14 steps of zooming into it; the color coding shows the number of iterations necessary to reach divergence. ^[38]

Julia set The Julia set is created similar to the Mandelbrot set ^[39]. It is also calculated in a complex plane by an iterative procedure according to Eq. 3.2 which represents a simple form of the Julia set for a quadratic polynomial. z and c are complex numbers. Unlike for the Mandelbrot set, c is a constant and z is varied

from $z = (-2 - 2i)$ to $z = (2 + 2i)$. The Mandelbrot set and the Julia set are strongly correlated. If the Julia set is calculated for a value of c that is part of the Mandelbrot set the resulting Julia set is connected, otherwise it is disconnected. For real values of c the Julia set is mirrored about a plane perpendicular to the complex plane containing the real axis otherwise it has a twofold axis perpendicular to the complex plane intersecting the origin. In Fig. 3.10 a Julia set was created with $c = (-0.52 + 0.58i)$ with a maximum number of iterations of 100 for $z = (-2 - 2i)$ to $z = (2 + 2i)$. The converging sets of z are drawn in black, all others in white.

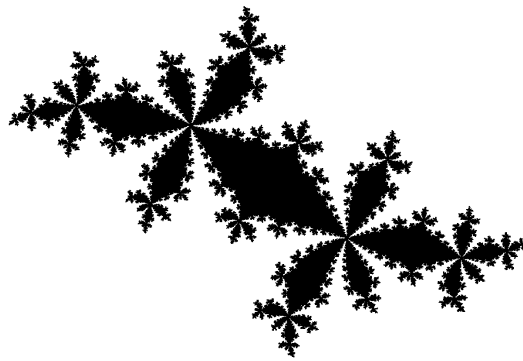


Figure 3.10: Julia set for a quadratic polynomial with $c = (-0.52 + 0.58i)$.

The twofold symmetry is changed to a sevenfold symmetry for a Julia set created with a polynomial of seventh order (Fig. 3.11). The colors indicate the number of iterations until divergence was reached.

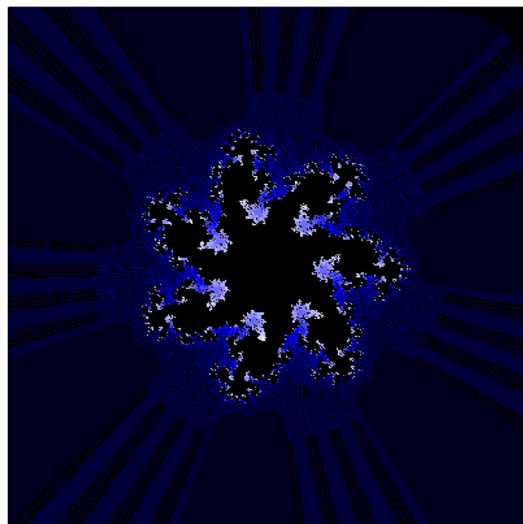


Figure 3.11: Julia set for a polynomial of seventh order with $c = (-0.74543 + 0.11301i)$.

For this Julia set the corresponding equation is

$$z_{n+1} = z_n^7 + c \quad (3.3)$$

with $c = (-0.74543 + 0.11301i)$ for $z = (-2 - 2i)$ to $z = (2 + 2i)$.

All those fractals are perfectly self-similar, that is subsections of the object are similar to the whole object. They are highly detailed even after multiple magnification. The dimension of those objects is not obvious, but in any case it exceeds its topological dimension.

3.1.2.2 Fractals in nature

Fractals can not only be constructed mathematically and be drawn by computers but they exist also in nature. However, the self-similarity is not given strictly anymore as the size of the building blocks is limited. Natural fractals can *e.g.* be found in the human body in the surface of the brain, pulmonary vessels or the systems of blood vessels. Other fractals in nature are mountains, clouds, river networks or the branches of trees. Two examples, a fern and a romanesco broccoli, are shown in Fig. 3.12. Parts of the whole are similar to the whole when the part is magnified, but only up to about 3–5-fold magnification.



(a) fern



(b) romanesco

Figure 3.12: Fractals in nature.

Very important examples for natural fractals are coastlines. A geographical problem that was solved with the help of the concept of the fractal dimensionality is the measurement of coastlines. At the beginning of the last century the scientist Lewis Fry Richardson was interested in determining the exact length of frontiers. By the attempt to show that the tendency to war in different regions is dependent on shapes, length and contiguities of these regions he discovered that he could not

find exact numbers for the length of frontiers. By examining maps and measuring the length of the frontiers with polygons, which have equal sides and have their corners on the frontier, he discovered that for the side of the polygon approaching zero the length of the frontier would increase to infinity. ^[40,41] The smaller the ruler the finer are the details that can be measured by it (see Fig. 3.13). And as there is in principle almost no limit in the size of the ruler there is also almost no limit in the length of the frontier. A ruler of a length of 500 km results in a length for the coastline of the main island of Great Britain of 2600 km, a ruler of 100 km in a length of 3800 km, a ruler of 54 km in a length of 5770 km and a ruler of 17 km in a length of 8640 km ^[42].



Figure 3.13: Coastline of the main island of Great Britain measured with rulers of different length.

When Richardson plotted the logarithm of the length g of the compass versus the logarithm of the total length $L(g)$ he found a linear correlation between the two, from which he derived the following empirical formula that was revived (in 1967) by Mandelbrot ^[43]

$$L(g) \propto g^{1-D} \quad (3.4)$$

with D being the dimension of the frontier. As the frontier can be described as a line the dimension must have a magnitude around 1, but the dimension D can also be a non-integer value larger than 1. Richardson analyzed the borders of several countries and determined the fractal dimensions of the west coastline of Great Britain to be $D = 1.25$, the land frontier of Germany (in 1900) $D = 1.15$ and $D = 1.02$ for the coast of South Africa. The more irregular the land frontier is the more the dimension deviates from 1. In contrast to the irregularly shaped land borders a circle has a

perimeter of finite length and can be measured exactly, which leads to a fractal dimension of its length of 1.

In Fig. 3.14 the log-log-representation of the scale versus the total length of the coastline of Great Britain and a circle with comparable perimeter are plotted according to the values found by Peitgen *et al.* [42]. The slope of the graph corresponds to the exponent of Eq. 3.4, which is $1 - D$. The slope of the red line is -0.36 whereas the slope of the blue line is -0.01 . This leads to a fractal dimension of the coastline of $D = 1.36$ and to a (fractal) dimension of the circle of $D = 1.01$. Although the west coast of Britain is more irregularly shaped than the east coast Richardson found a lower value for the fractal dimension of the west coast than Peitgen *et al.* found for the whole island (1.25 compared to 1.36). This might be due to the fact that Richardson had an older map which might have shown less details than the map used by Peitgen *et al.*. A less detailed map leads to a lower value for D .

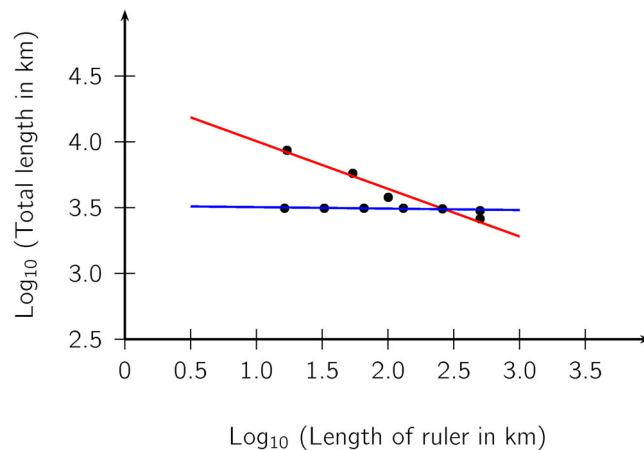


Figure 3.14: log-log-plot for the coastline of Britain (red) and a circle (blue).

Another possibility instead of determining the length of the coastline by measuring it with a compass is to evaluate the length with a box-counting method.

In this approach the structure of which the dimension is to be determined is covered by a grid built up by boxes.

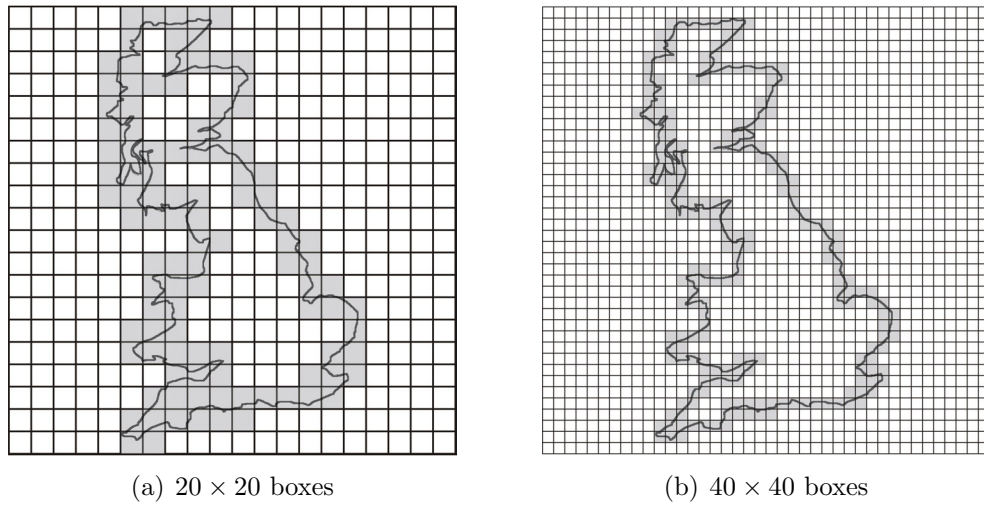


Figure 3.15: Coastline of Britain on two different grids; the boxes containing the coastline are plotted in gray.

The number of boxes $N(\varepsilon)$ which contain *e.g.* the coastline of Great Britain are counted and set into relation to the length ε of the boxes. The fractal dimension determined by this method is called box-counting dimension or Hausdorff dimension (sometimes Hausdorff-Besicovitch dimension) d^f and is defined in the limit of infinitesimal small boxes ^[44,45].

$$d^f = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \frac{1}{\varepsilon}} \quad (3.5)$$

Fig. 3.15(a) shows the coastline of Great Britain covered by a grid of 20×20 boxes. The gray boxes contain the coastline. For the 20×20 grid this number is 105, whereas the 40×40 grid in Fig. 3.15(b) has 215 gray boxes that cover the coastline. For Fig. 3.15(a) $d^f = \frac{\log 105}{\log 20} = 1.55$, whereas for Fig. 3.15(b) the fractal dimension is $d^f = \frac{\log 215}{\log 40} = 1.46$. One can see that for the finer grid with smaller boxes the fractal dimension decreases and approaches slowly the value of $D = 1.36$ that was found with the method of Richardson.

3.1.3 The dimension of fractals

According to Eq. 3.5 the dimension of the fractals that were introduced in the previous sections can be calculated easily (see Fig. 3.16). For the Cantor dust, of which the construction started from a line and ended in a set of points, the fractal dimension can be calculated to be 0.6309. The Sierpinski triangle and the Sierpinski carpet started from a plane from which subsequently parts were removed and thus their fractal dimensions are 1.5850 and 1.8928, respectively. The Koch curve was

constructed from a line which was modified such that its dimension exceeds 1 and is 1.2619. The Peano curve also started from a line and is arranged such that it even completely fills a plane and has a dimension of 2.

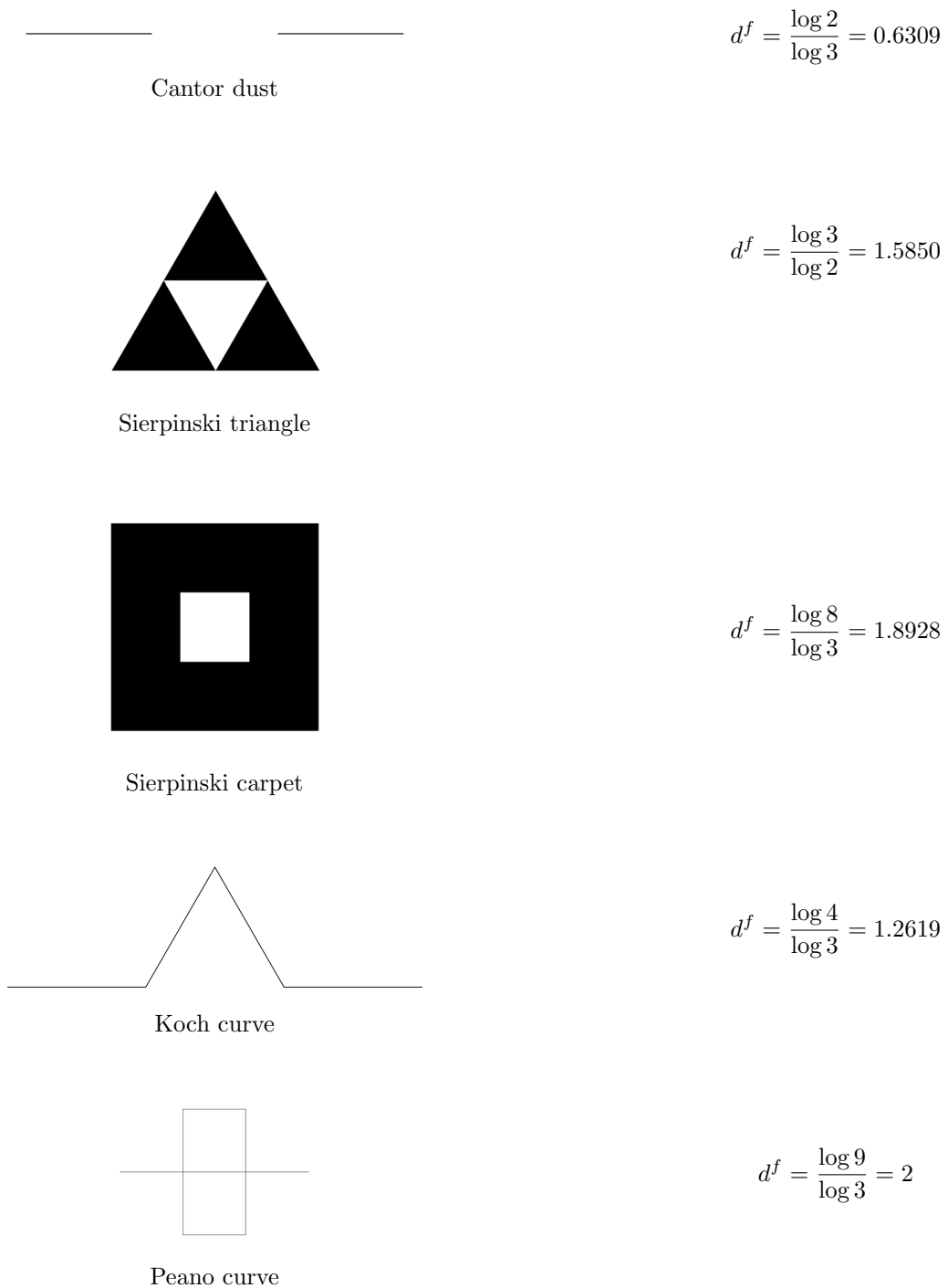


Figure 3.16: Hausdorff dimension of some mathematical fractals.

3.1.4 The fractal dimension of the residual density

The distribution of the residual density in the unit cell can also be considered as an irregularly shaped line (if it is drawn in a 2-dimensional plot) or plane (in the 3-dimensional case) and thus its quantification by its fractal dimension can give a number that is on an absolute scale and allows a comparison of residual densities of different structures. In the hypothetical case where no noise is present in the data and the model describes perfectly the electron density ($F_{obs} = F_{calc}$) which is present in the unit cell the iso-surface of $\rho_0 = 0 \text{ e}\text{\AA}^{-3}$ completely fills the unit cell. In this case, the dimension of the residual density value $\rho_0 = 0 \text{ e}\text{\AA}^{-3}$ is 3. This can also be calculated easily with the box-counting dimension: if the unit cell with the residual density is covered by a $10 \times 10 \times 10$ grid then each of the resulting boxes contains the zero-surface. This leads to a number of filled boxes $N(\varepsilon)$ of 1000 and a box size ε of $\frac{1}{10}$. According to Eq. 3.5 it follows that $d^f = \frac{\log 1000}{\log 10} = 3$. If no model errors but only noise was present in the data the extension of the zero-residual density decreases and its dimension is reduced. Instead, positive and negative residual density appears. Fig. 3.17 shows the fractal dimension distribution for a simulated data set with limited experimental resolution where only noise is present in the data. As was expected, the shape of the distribution is that of a parabola. The upper limit for the fractal dimension distribution is 3 for a space-filling distribution of the zero residual density and this value decreases to a non-integer value for any deviation of the absolute ideal case, which corresponds to the real case. The height of the graph denotes the featurelessness (the higher the maximum at $\rho_0 = 0 \text{ e}\text{\AA}^{-3}$ the less features are in the distribution of the residual density) and the width of the base line denotes the flatness ($\Delta\rho_0 = \rho_{0,max} - \rho_{0,min}$).

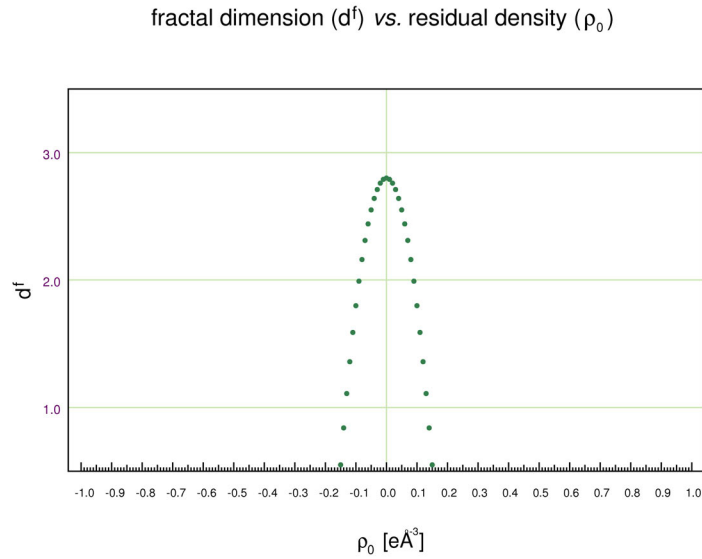


Figure 3.17: Ideal parabolic shape for the fractal dimension distribution if only Gaussian noise is present.

3.2 Development of new descriptors

For the characterization of the residual density the following descriptors were developed:

3.2.1 Gross residual density and gross residual electrons

To quantify the total amount of residual density the gross residual density and gross residual electrons were introduced.

The gross residual density ρ_{gross} is the integral over all absolute values ρ_0 of the residual density.

$$\rho_{gross} = \frac{1}{2V} \int_V |\rho_0(xyz)| dV \quad (3.6)$$

For a residual density grid this turns into a summation of the moduli of all N residual density values ρ_0 from the grid.

$$\rho_{gross} = \frac{1}{2N} \sum_{i=1}^N |\rho_0(i)| \quad (3.7)$$

The factor $\frac{1}{2}$ corrects for the fact that one misplaced electron causes a negative

residual electron in the region where it is wrongly put and one positive residual electron where it actually should be.

The gross residual electrons e_{gross} are simply calculated from the gross residual density by taking the volume under consideration into account

$$e_{gross} = V \cdot \rho_{gross} = \frac{V}{2N} \sum_{i=1}^N |\rho_0(i)| \quad (3.8)$$

The gross residual electrons reflect the noise contained in the data, inadequate data processing and density model errors. It showed to be independent of the grid size but it can be artificially reduced when the experimental resolution is truncated. For the best model e_{gross} should be reduced to a minimum although it would never reach a value of 0e due to the noise that will remain present in the data.

3.2.2 Net residual density and net residual electrons

The net residual density ρ_{net} is the integral over all residual density values. In contrast to the gross residual density the values are taken with their corresponding signs.

$$\rho_{net} = \frac{1}{V} \int_V \rho_0(xyz) dV \quad (3.9)$$

For a grid this corresponds to a summation over all residual density values:

$$\rho_{net} = \frac{1}{N} \sum_{i=1}^N \rho_0(i) \quad (3.10)$$

By taking again the volume into account the net residual electrons e_{net} are obtained.

$$e_{net} = V \cdot \rho_{net} = \frac{V}{N} \sum_{i=1}^N \rho_0(i) \quad (3.11)$$

In contrast to the gross residual electrons a value of 0e should be obtained for e_{net} if the volume under consideration is the volume of the unit cell (for a detailed calculation of this value see appendix A.1). F_{000} is the structure factor resulting from a diffraction of the electron cloud at a Bragg angle of $\theta = 0^\circ$, but unfortunately it cannot be measured as it is obscured by the primary beam, which is hidden by the beam stop. F_{000} corresponds to the number of electrons in the unit cell and thus the mean contribution of all other structure factors to the integration of the electron density over the unit cell is 0. The scale factor is calculated such that

the measured structure factors are scaled to the observed structure factors, *i.e.* the number of electrons that are described by both F_{000} are equal. The integration over their difference, which is the residual density, must be zero for the unit cell. Thus, e_{net} , which (in principle) corresponds to the integration over the residual density, must also be 0 for the unit cell. If F_{000} could be measured exactly (also without contribution of noise) the scale factor that scales the calculated structure factors F_{calc} to the observed structure factors F_{obs} could be calculated instead of being refined and e_{net} would exactly describe the noise contained in the data, which does not average out. If e_{net} is calculated for a volume smaller than the unit cell it also should be zero if no noise or model errors are present but as this is not the case it will have a non-zero value that can be positive or negative, which quantifies the excess or missing number of electrons in this volume.

As indeed e_{net} was always zero within the exactness of the given values (four decimal places) its value will not be given in all following tables where the residual density descriptors for the whole unit cell are shown.

For a given grid of residual density there are some limitations on ρ_{gross} and e_{gross} , and ρ_{net} and e_{net} , respectively, following from Eqs. 3.7, 3.8 and Eqs. 3.10, 3.11: $-\rho_{gross} \leq \frac{1}{2}\rho_{net} \leq \rho_{gross}$ and $-e_{gross} \leq \frac{1}{2}e_{net} \leq e_{gross}$.

3.3 Implementation of the new descriptors

3.3.1 jnk2RDA

A software called 'jnk2RDA' was written to perform the analysis of the residual density grids. It calculates the fractal dimension distribution from $\rho_{0,min}$ to $\rho_{0,max}$ and plots a graphical representation into a PostScript file. It also prints out the fractal dimension at $\rho_0 = 0 \text{ e}\text{\AA}^{-3}$, the number of gross residual electrons e_{gross} and the net residual electrons e_{net} as well as the values for $\rho_{0,min}$, $\rho_{0,max}$ and their difference $\Delta\rho_0$. Moreover, the residual density values for which $d^f(\rho_0) = 2$ is valid are calculated from a linear interpolation. Also some information on the grid size (the number of n_x , n_y , n_z), the multipole model that was used for the refinement (MODEL), the two models that were used for the calculation of the Fourier file xd.fou (FOUR) and the XDFOUR section from the xd.mas file are printed.

jnk2RDA was written in Perl and can be used under the operating systems Linux and Windows. It processes residual density grid files from XDFOUR, XDFFT, MoPro [46,47] (if the file is given in XPLOR format) and grid files written by Tonto [48]. When the program is called it first asks for the format of the grid file and then for the name. The default option is a grid file from XDFOUR with the name 'xd_fou.grd',

which can be chosen by twice hitting the ENTER button. If the input file is in XDFOUR format also some information from the xd.mas file is read in, assumed that it is in the same folder as the grid file. These are the unit cell, the title of the project and the whole XDFOUR section. If no xd.mas file is found the program prints a corresponding message.

From the XDFOUR grid file the parameters n_x , n_y and n_z are read in. Furthermore, the number of objects is read in. If this number is 0 it is assumed that the grid comprises the whole unit cell (calculated with the option “cryst” for the limits from 0 to 1 in x -, y -, and z -direction), otherwise this gives the number of atoms in the asymmetric unit. In this case it is assumed that the grid was calculated only for a part of the unit cell (*i.e.* the grid was calculated with the option “3-points”). This information is important for the calculation of the volume under consideration. If the grid is for a part of the unit cell its volume must be calculated from the size of the cuboid which is given in the XDFOUR file. Otherwise, if the grid was calculated for the unit cell, the values at the boundary planes repeat and must not be counted twice. All this is taken into account by jnk2RDA.

For a MoPro grid file the unit cell dimensions are read in from the file. The six numbers in line 6 are the cell axes a , b and c and the cell angles α , β and γ . The grid is assumed to be calculated for the whole unit cell. The number of grid points along the axes is read in from line 5 of the grid file. There are nine numbers in this line where the third number is n_z , the sixth number is n_y and the ninth number is n_x .

If an input file from Tonto is chosen, the user has the choice between reading in two grid files or one. When two files are read in their values are subtracted from each other which allows for the determination of *e.g.* difference densities. If only one file is entered it is assumed that this file contains residual density values and it is processed accordingly.

Finally, the residual density values (for a XD file six values per line are assumed) are read in and evaluated. In the present XD version the grid values are calculated in planes along the z -direction, with the x -direction varying fastest. jnk2RDA can handle 2-dimensional as well as 3-dimensional grids. Whether the residual density has been calculated for a plane or a cuboid is determined by the value of n_z : if n_z is 1 the dimension is 2, for all values of n_z larger than 1 the dimension is 3. For the evaluation of the residual density values several algorithms were investigated which are presented in the following.

3.3.2 Box counting algorithm

In the box counting algorithm, which is a standard procedure well-known from the literature^[44], the three-dimensional residual density is covered by cuboids. Those are formed by reducing eight neighboring grid values forming the corners of a cuboid to a single value, which could be -1 , 0 or 1 . A value of -1 is assigned when all contributing values are smaller, and $+1$ when they are larger than the level of residual density under consideration, respectively. A value of 0 is assigned when at least one of the eight grid points in this cuboid corresponds to the value under consideration or when from two neighboring values in x -, y -, and z -direction, respectively, one is smaller and the other one is larger than the value of the level under consideration. This is done for all residual density levels from the lowest to the highest residual density value in steps of $0.01 \text{ e}\text{\AA}^{-3}$. In the two-dimensional case quadrangles are taken instead of cuboids. An example for how the number of boxes which contain the level of residual density under consideration is determined is shown in Fig. 3.18. Fig. 3.18(a) shows the distribution of the zero line of the residual density in a plane in the molecule. The residual density grid was calculated by a Fourier back transformation with the program XDFOUR. The 25×25 grid was plotted with the program MAPVIEW which is part of the XD2006 program package. Fig. 3.18(b) shows the same plane with an overlay of grid lines. 25×25 grid points are thus reduced to 24×24 boxes. In Fig. 3.18(c) the boxes containing the zero line are plotted in gray. Counting the gray boxes gives $N(\varepsilon) = 504$ resulting in a fractal dimension according to Eq. 3.5 of $d^f(0) = \frac{\log 504}{\log 24} = 1.9580$. In Fig. 3.18(d) the grid resolution is increased to 50×50 grid points which reduce to 49×49 boxes. It can easily be seen that increasing the grid resolution decreases the relative number of gray boxes. The fractal dimension thus reduces to $d^f(0) = \frac{\log 1759}{\log 49} = 1.9201$. Increasing the grid resolution to a 50×50 grid results in a relative decrease in the number of boxes that contain at least one point of the zero line.

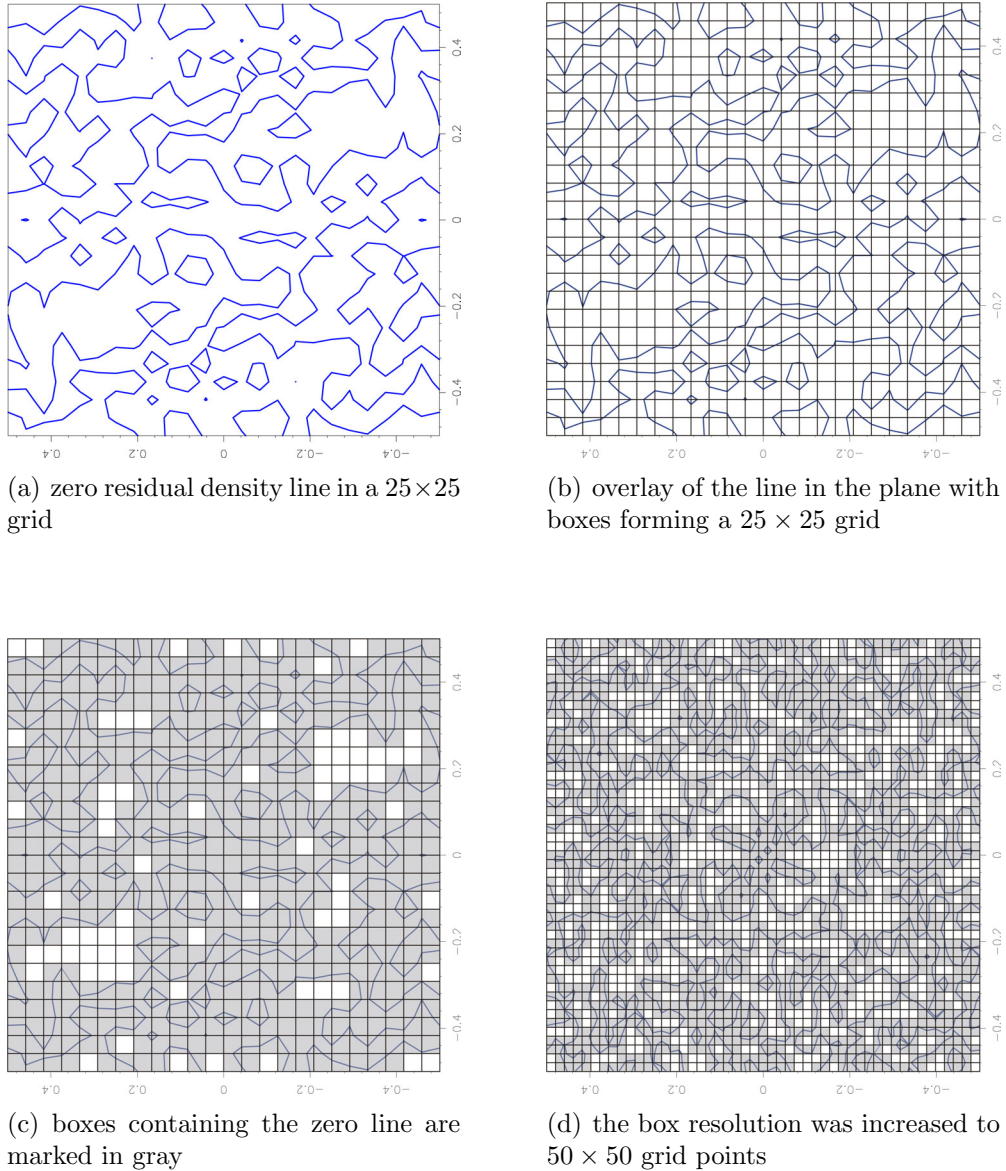


Figure 3.18: 2-dimensional residual density grids with different grid sizes for the exemplification of the box counting algorithm.

Although the box-counting algorithm is a standard technique for the determination of the fractal dimension it turned out to be not very efficient for our purposes.

Thus, a modification of the box-counting algorithm was developed to meet the requirements with respect to accuracy and efficiency. This algorithm was called line-counting algorithm and is described in detail in the following section.

3.3.3 Line-counting algorithm

In the line-counting algorithm consecutive values in the residual density grid file are compared. If the value under consideration lies between them the counter for this

value is increased by one. This procedure is performed in x -, y - and z -direction for all existing residual density levels in steps of $0.01 \text{ e}\text{\AA}^{-3}$. Fig. 3.19 demonstrates this procedure for a 2-dimensional plot of the residual density. As an example the zero residual density value (blue line in Fig. 3.19) is evaluated. The black lines connect each two consecutive values which are separated by the zero residual density line and their number is counted. In the given example for a 25×25 grid this gives a value for $N(\varepsilon)$ of 574. $\frac{1}{\varepsilon}$ is calculated from the number of grid points n_x and n_y with

$$\frac{1}{\varepsilon} = \log \sqrt{2(n_x - 1)(n_y - 1)} \quad (3.12)$$

for a 2-dimensional grid. According to Eq. 3.5 this yields a fractal dimension of $d^f(0) = \frac{\log 574}{\log \sqrt{24 \cdot 24 + 24 \cdot 24}} = 1.8024$. For the line-counting algorithm the denominator is calculated by taking the logarithm of the square root of all possible lines connecting two consecutive values in the 2-dimensional case, and taking the logarithm of the cubic root in case of a 3-dimensional grid. The value for $d^f(0)$, calculated with the line-counting algorithm, is found to be systematically decreased by about 0.2 in comparison to the box-counting algorithm. That different methods for calculating the fractal dimension result in slightly different values for d^f is well-known from the literature [49].

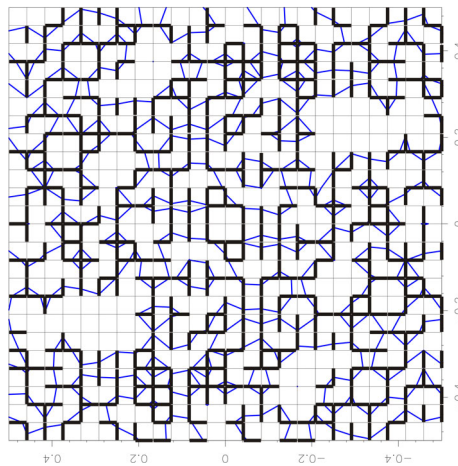


Figure 3.19: Line-counting algorithm on a 25×25 grid.

3.3.4 Optimization of the program jnk2RDA

The first version of jnk2RDA used the box-counting algorithm and stored all the residual density values in memory before evaluating them. This was very CPU time and memory consuming. The analysis of a $100 \times 100 \times 100$ grid took about 25 min

on an AMD Dual Opteron (64 Bit) processor with 2.4 GHz and 4 GB RAM and a $200 \times 200 \times 200$ grid could not even be evaluated because the program was aborted by the operating system due to memory overflow.

At this stage the line-counting algorithm was used instead of the box-counting algorithm. This sped up the processing of a $100 \times 100 \times 100$ grid by a factor of almost 200 from 25 min to 8 s. The $200 \times 200 \times 200$ grid could now be analyzed within 72 s. But a $400 \times 400 \times 400$ grid was still too big to be evaluated with jnk2RDA. As the value for $d^f(0)$ still was not converged further optimization was required. So far, all values of the grid were stored simultaneously in memory.

A new version was written that now processes the data when they are read in. For the evaluation of the grid points in x -direction only the two consecutive values have to be stored in memory at the same time. For evaluation in y -direction only two consecutive blocks of n_x values each are stored in the memory and for evaluation in z -direction two consecutive blocks of $n_x \times n_y$ values are stored at the same time. So a 3-dimensional residual density grid is reduced to a 2-dimensional memory problem. For the above-mentioned $400 \times 400 \times 400$ grid this means that instead of 64 000 000 only 320 000 numbers have to be stored simultaneously in memory. As the residual density values are evaluated in a stepwidth of $0.01 \text{ e}\text{\AA}^{-3}$ they are converted to integer values by multiplying them with the reciprocal of the stepwidth (100 in this case) which reduces the number of characters that have to be stored. By conversion into integers the problem of internal storage of floating point numbers is bypassed, too. The largest grid that was evaluated with jnk2RDA had a size of $1280 \times 1280 \times 1280$ and it took about 3 h to analyze it. For such large grids the limiting factor is the size of the file (the $1280 \times 1280 \times 1280$ grid file had a size of 36 GB) and not the CPU time or memory demand of jnk2RDA. For large grids XDFOUR needs much longer for the Fourier transformation than jnk2RDA needs to evaluate the data.

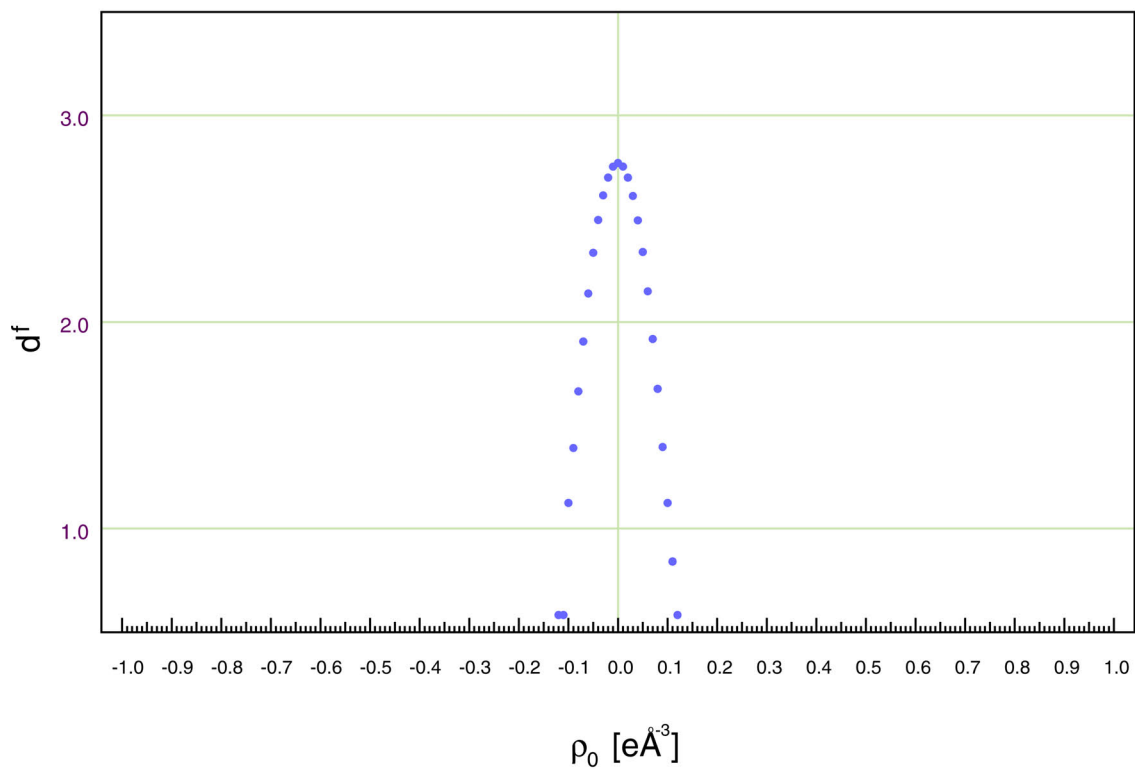
3.3.5 Example output file from jnk2RDA

Fig. 3.20 shows a typical output file from jnk2RDA. The name of the PostScript file is derived from the input grid file by just concatenating the name of the grid file with the ending “_jnk2rda.ps”. If the xd.mas file is present in the folder with the grid file some information on the data is printed in the .ps file like the title of the project on top of the page, and in the left gray box the multipole model that was used for the refinement (MODEL), the two models for which the structure factors are calculated (FOUR) and the XDFOUR section. In the right gray box the values for $d^f(0)$, $\rho_{0,min}(d = 2)$, $\rho_{0,max}(d = 2)$, $\rho_{0,min}$, $\rho_{0,max}$, $\Delta\rho_0$ and the grid size in n_x , n_y and n_z are printed. Below this box the values for e_{gross} and e_{net} are given. In

the bottom line the name of the program (jnk2RDA), the path to the grid file, the name of the grid file and the date when the grid was analyzed are printed.

The fractal dimension distribution in this example file shows an ideal parabolic shape that is obtained if only Gaussian noise is present in the data.

STERN

fractal dimension (d^f) vs. residual density (ρ_0)

```

MODEL model 4 2 1 0
FOUR fmod1 4 2 0 0 fmod2 -2 2 0 0

SELECT *fobs *fmod1 fmod2 print snlmin 0. snlmax 2.
GRID 3-points perp *cryst
LIMITS xmin 0.0 xmax 1.0 nx 50
LIMITS ymin 0.0 ymax 1.0 ny 50
LIMITS zmin 0.0 zmax 1.0 nz 50

```

```

d^f(0) = 2.7693
rho_min(d=2) = -0.0659 eA^-3
rho_max(d=2) = 0.0664 eA^-3
nx=50      rho_min: -0.1214 eA^-3
ny=50      rho_max: 0.1256 eA^-3
nz=50      delta rho: 0.2470 eA^-3

```

e_gross: 8.3803 e

e_net: 0.0000 e

jnk2RDA.pl

Q:\grid_check\fraktale_Dimension\test-daten\trimid\invent_HM\1.14\per\gaussnoise\p1_0.222\cycle0

xd_fou.grd

31.07.2008, 01:54:41

Figure 3.20: Example output file from jnk2RDA.

3.4 The prediction of the number of gross residual electrons in the whole unit cell

The number of gross residual electrons in the unit cell can under certain assumptions be calculated from the modulus squared of the Fourier-coefficients of the residual density by the following equation:

$$e_{gross} = \frac{1}{\sqrt{2\pi}} \sqrt{\sum_{hkl} |\Delta F|^2} \quad (3.13)$$

For a detailed derivation of this equation see appendix [A.2](#).

3.5 Results and Discussion

3.5.1 Application of new descriptors to simulated data

3.5.1.1 Influence of the crystallographic resolution

Noise-free ideal simulated data were generated from the model for testing the pure influence of the variation of different parameters on the residual density. Those data are ideal and noise-free in the sense that $F_{obs} = F_{calc}$ for all data up to $\sin \theta / \lambda = 1.14 \text{ \AA}^{-1}$. For this purpose, XDLSM was run to generate an `xd.fco` file which contains F_{obs}^2 and F_{calc}^2 . The F_{calc}^2 were extracted from the `xd.fco` file with the program `fco2hkl` and pasted into the `xd.hkl` file. With those ideal F_{obs}^2 ten cycles of a least-squares refinement of all parameters against F^2 with no $\sigma(I)$ cutoff ($I > 0\sigma(I)$) were performed with XDLSM. The final R -factor that was achieved by applying this procedure was 0.01%. Although the option “cycle 0” in XD should only perform a structure factor calculation the F_{obs} and the F_{calc} are not identical but can deviate by 0.01, which could come from rounding errors and leads to the observed numerical error. To investigate the dependence of $d^f(0)$ on the crystallographic resolution several data files with ideal structure factor sets were generated for different resolutions. This was done by calculating initial reflections with an intensity of 1.0 and a corresponding σ of 1.0 with the program `invent_hkl` (see section 5) up to the desired resolution. With these data XDLSM was run with the option “cycle 0” to calculate the ideal structure factors for the input multipole model. Then the F_{calc} were transferred to the `xd.hkl` file with `fco2hkl`. As the structure factors that were used as F_{obs} were calculated from the F_{calc} the resulting R -value from a subsequent least-squares refinement with XDLSM (10 cycles) was not more than

0.01 %. Residual density grid files ($50 \times 50 \times 50$ grid points for the whole unit cell) were calculated with XDFOUR for the calculated resolution range and the data were analyzed with jnk2RDA. The results are given in Table 3.1.

$\sin \theta/\lambda$ (\AA^{-1})	d (\AA)	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
0.80	0.63	2.7093	0.1058	-0.00	0.00	0.00
0.90	0.56	2.7383	0.1475	-0.00	0.00	0.00
1.00	0.50	2.7647	0.1970	-0.00	0.00	0.00
1.10	0.45	2.7901	0.2986	-0.00	0.00	0.00
1.14	0.44	2.7956	0.3361	-0.00	0.00	0.00
1.20	0.42	2.8083	0.4096	-0.01	0.01	0.02
1.30	0.38	2.8193	0.5802	-0.01	0.01	0.02
1.40	0.36	2.8293	0.7970	-0.01	0.01	0.02
1.50	0.33	2.8365	1.0307	-0.02	0.01	0.03

Table 3.1: Residual density descriptors applied to simulated data for different resolutions for $S(\text{N}^t\text{Bu})_3$.

From Table 3.1 it can be seen that $d^f(0)$ increases steadily with increasing resolution and e_{gross} as well as $\Delta\rho_0$ increase, too. Although the residual density values for the lower resolution are all smaller than $\pm 0.01 e\text{\AA}^{-3}$ $d^f(0)$ does not reach the value of 3. Instead, even though the residual density values for higher resolution data increase up to about $\pm 0.02 e\text{\AA}^{-3}$ the fractal dimension $d^f(0)$ reaches larger values than with the “flatter” residual density. With increasing resolution of the data the number of reflections increases drastically (the reflection file for 0.8\AA^{-1} contains 6795 entries, for 0.9\AA^{-1} the number is already 9654 and the file with 1.5\AA^{-1} comprises 38980 reflections) and thus the aforementioned numerical errors introduced by XD accumulate. This causes the increase in the number of e_{gross} and $\Delta\rho_0$. With increasing resolution it occurs more often that two neighboring residual density values change their sign and it is the frequency of the change in sign of two neighboring values that contributes to the fractal dimension. Thus the fractal dimension increases for increasing resolution and constant grid size. This also explains why the theoretical value of 3 for $d^f(0)$ for residual density of $0 e\text{\AA}^{-3}$ all over the unit cell is never reached. This would require all consecutive values to have opposite sign and is quite unlikely to happen. The resolution of the experimental data was 1.14\AA^{-1} and the fractal dimension for this resolution was calculated to be 2.7956. This is the maximum value that could be reached for $d^f(0)$ if no noise and no model errors were present in the data for the given experimental resolution and residual density grid size.

3.5.1.2 Influence of Gaussian noise

Noise in the data is unavoidable due to counting statistics, cosmological radiation, thermal or shot noise generating unwanted currents and voltages, thus, its influence on the residual density, the fractal dimension, e_{gross} and on the flatness was investigated. As the noise in the data can to a good approximation be assumed to be Gaussian distributed [23], again ideal data were calculated and Gaussian random numbers which should represent the noise were added to the ideal intensities. The ideal data were generated for a resolution of 1.14 \AA^{-1} (which corresponds to the experimental resolution) according to the procedure that was already described in the previous section. The noise (I_{hkl}^{error}) was added to the ideal intensities (I_{hkl}^{ideal}) by generating Gaussian random numbers with the polar form of the Box-Muller transform [50] which results in the new noise-containing intensities I_{hkl}^{noise} .

$$I_{hkl}^{noise} = I_{hkl}^{ideal} + I_{hkl}^{error} \quad (3.14)$$

The error intensity is distributed according to a Gaussian probability density function (RandomGN).

$$I_{hkl}^{error} = p_1 \sqrt{I_{hkl}^{ideal}} \text{RandomGN}(x) \quad (3.15)$$

with

$$\begin{aligned} \text{RandomGN}(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ \sigma &= 1 \\ \mu &= 0 \end{aligned}$$

p_1 is an adjustable noise control parameter with $0 \leq p_1 \leq 1$.

I_{hkl}^{error} assumes positive and negative values with the same probability. Several hkl files with different noise levels (controlled by p_1 , see Eq. 3.15) were created with factor_int_sig (see section 5). The standard deviation of the intensities was set to 1.0. A structure factor calculation with XDLSM (“cycle 0” with an exclusion of negative intensities) was performed and the residual density grid files were calculated with XDFOUR (grid size $50 \times 50 \times 50$ for the whole unit cell). The results were analyzed with jnk2RDA and are shown in Table 3.2 and Fig. 3.21.

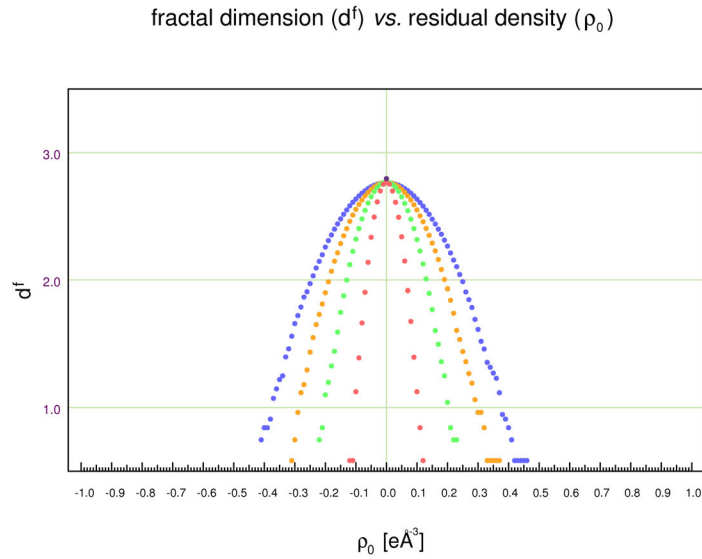


Figure 3.21: Ideal parabolic shape for the fractal dimension distribution in the presence of Gaussian noise and absence of model errors.

p_1	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
0.000	2.7956	0.3361	-0.00	0.00	0.00
0.222	2.7693	8.3803	-0.12	0.12	0.24
0.444	2.7678	16.4095	-0.22	0.23	0.45
0.666	2.7663	23.8102	-0.31	0.37	0.68
0.888	2.7647	30.9192	-0.41	0.46	0.87

Table 3.2: Residual density descriptors applied to simulated data for different amount of Gaussian noise for $S(N^tBu)_3$.

As can be seen from Table 3.2 the fractal dimension decreases with increasing noise. It does not decrease linearly as one could expect but it drops down from $d^f(0) = 2.7956$ to 2.7693 when noise is introduced and then remains quite stable at this value. The decrease of $d^f(0)$ when p_1 is increased from 0.000 to 0.222 results from the reduction of the space the zero residual density can fill. As a consequence $d^f(0)$ is decreased. A further increase of noise with no adjustment of the parameters (“cycle 0” in XDLSM) only increases the values of the residual density but does not change the spatial distribution. Thus a further increase in p_1 does not decrease $d^f(0)$ with the same ratio. With more noise present in the data also the flatness decreases as can be seen from the increase of $\Delta\rho_0$. The ratio of $\rho_{0,min}$ to $\rho_{0,max}$ is not balanced, $\rho_{0,max}$ tends to be slightly larger than $\rho_{0,min}$. This results from the criterion for observed intensities which includes all structure factors with a positive

intensity but excludes the negative ones. As the introduced noise can have positive as well as negative values this leads to some negative intensities. The corresponding structure factors F_{obs} are omitted from the calculation of the residual density and this leads to a bias towards the positive residual density values. The gross residual electrons for $p_1 = 0.222$ ($e_{gross} = 8.3803 e$) are of about the same magnitude as for the experiment ($e_{gross} = 8.3961 e$), which means that also the noise (if only Gaussian noise is present in the data) is of about the same magnitude. Thus, the value of the fractal dimension for $p_1 = 0.222$ ($d^f(0) = 2.7693$) can be seen as the upper limit for $d^f(0)$ for experimental data on $S(N^tBu)_3$ if only noise remained after modeling the data.

3.5.1.3 Analysis of the impact of change of model parameters on the residual density

To investigate the impact of several parameters of the multipole model on the residual density ideal noise-free data from the fully refined model of $S(N^tBu)_3$ were calculated for a resolution of $\sin \theta/\lambda = 1.14 \text{ \AA}^{-1}$ and some of the multipole parameters were changed to values deliberately deviating from the original ones. These parameters were the scale factor, the anisotropic displacement parameters U_{ij} of the sulfur atom, the expansion/contraction parameters of the spherical valence (κ) and the aspherical deformation density (κ') of the sulfur atom and the anomalous dispersion values $\Delta f'$ and $\Delta f''$ of the sulfur atom. The original values for the respective parameters are given in Table 3.3, the starting distribution of the residual density is shown in Fig. 3.22(a) for the main plane of the molecule and in Fig. 3.22(b) for a $50 \times 50 \times 50$ grid over the whole unit cell.

Ideal data were generated with the program `fco2hkl` (see section 5) by replacing the F_{obs}^2 with their corresponding F_{calc}^2 from the final refinement and writing them into a new reflection file `xd.hkl`.

For all subsequent parameter manipulations the parameters were all changed independent of each other and no refinement was performed (“cycle 0” in XDLSM). Only positive intensities were used for the refinement ($I/\sigma(I) > 0$). Two Fourier transformations with XDFOUR were applied to calculate one grid with 50×50 data points in a plane containing the molecule and another grid of $50 \times 50 \times 50$ data points covering the complete unit cell. The figures on the left hand side show the residual density in the plane containing the molecule and were drawn with XDGRAPH. The figures on the right hand side show the fractal dimension distribution which was produced with `jnk2RDA` on the $50 \times 50 \times 50$ residual density grid for the unit cell.

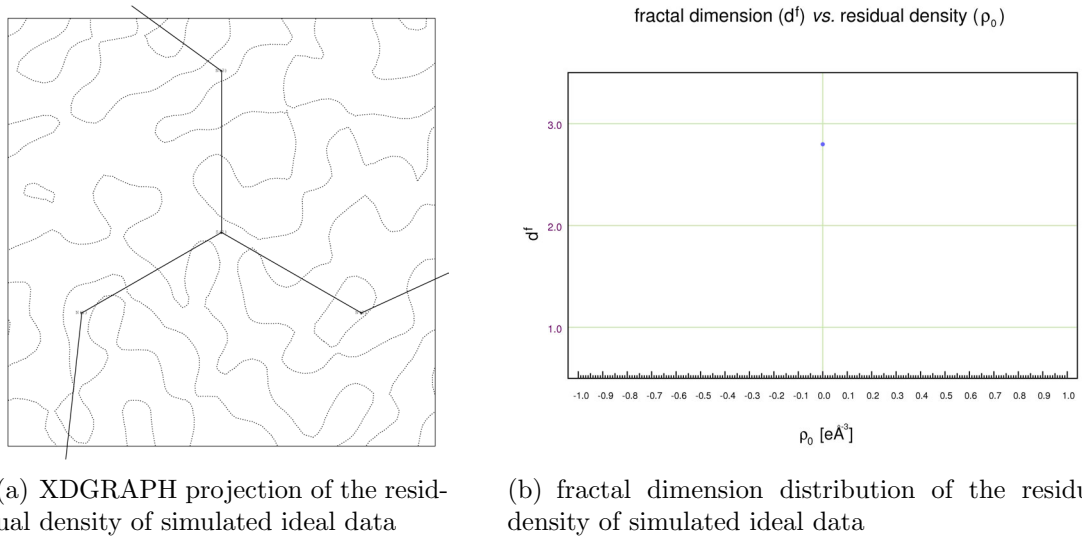


Figure 3.22: Residual density after the multipole refinement of $S(N^tBu)_3$ on ideal, noise-free data for $\sin \theta/\lambda \leq 1.14 \text{ \AA}^{-1}$; (a) gray dotted lines: zero residual density.

$d^f(0)$		2.7972		
e_{gross}	(e)	0.3136		
$\rho_{0,min}$	($e\text{\AA}^{-3}$)	-0.00		
$\rho_{0,max}$	($e\text{\AA}^{-3}$)	0.00		
$\Delta\rho_0$	($e\text{\AA}^{-3}$)	0.00		
κ		1.088207		
κ'		1.192135		
U_{ij}	(\AA^2)	U_{11} : 0.011311	U_{22} : 0.012366	U_{33} : 0.024020
		U_{12} : -0.005632	U_{13} : 0.000134	U_{23} : -0.006742
$\Delta f'$		0.1246		
$\Delta f''$		0.1234		
Scale factor k		0.100000E+01		

Table 3.3: Original values of S for simulated ideal data of $S(N^tBu)_3$.

As the data are ideal in the sense that the F_{obs} correspond to the F_{calc} the residual density is flat and featureless as expected. $d^f(0)$ is at its optimal value for the given resolution if no noise was present and e_{gross} takes its minimum value. That the value of e_{gross} is not equal to zero results from the fact that despite the option “cycle 0”

XD calculates new intensities F_{calc}^2 that can slightly deviate (by 0.01) from the F_{obs}^2 . The flatness is zero as the residual density values do not exceed $\pm 0.01 \text{ e}\text{\AA}^{-3}$.

Manipulation of the scale factor k

As the measured structure factors F_{obs} are on an arbitrary scale whereas the calculated structure factors F_{calc} are calculated such that F_{000} corresponds to the total number of electrons in the unit cell they must be brought on the same scale to be comparable to each other. The scale factor k is determined during the refinement and scales the F_{calc} to the F_{obs} .

$$F_{obs} = k \cdot F_{calc} \quad (3.16)$$

As the data are ideal, *i.e.* the F_{obs} correspond to the F_{calc} , the starting value for the scale factor is 1. An increase in the scale factor scales the F_{calc} higher than they should be and thus increases the negative residual density as $F_{obs} - F_{calc}$ decreases (*i.e.* gets more negative). Fig. 3.23(a) shows the main plane of the molecule after an increase in the scale factor by 1%, in Fig. 3.23(b) the corresponding fractal dimension distribution is shown. From both figures it can be seen that the negative residual density increased which occurs mainly at the atomic positions.

A decrease in the scale factor results in an increase in the positive residual density as F_{obs} is scaled to higher values which leads to positive values for $F_{obs} - F_{calc}$. The values for the residual density descriptors for both an increase as well as a decrease of the scale factor are given in Table 3.4.

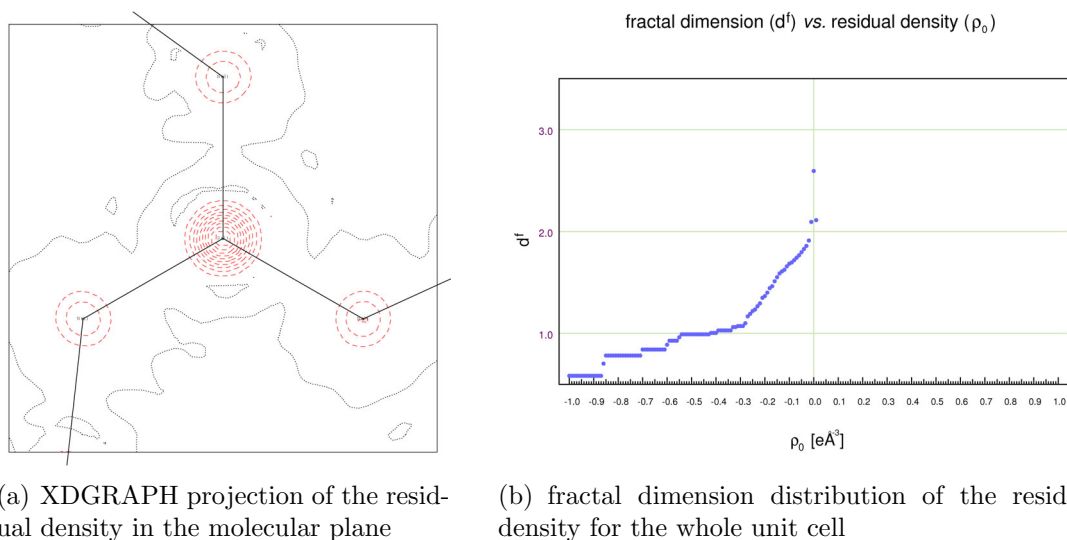


Figure 3.23: Residual density after an increase in the scale factor k in $\text{S}(\text{N}^t\text{Bu})_3$ by 1 %; (a) red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

	scale factor	$d^f(0)$	e_{gross}	$\rho_{0,min}$	$\rho_{0,max}$	$\Delta\rho_0$
			(e)	($\text{e}\text{\AA}^{-3}$)	($\text{e}\text{\AA}^{-3}$)	($\text{e}\text{\AA}^{-3}$)
-1 %	0.99000	2.5947	1.7257	-0.02	1.02	1.04
orig	1.00000	2.7972	0.3136	-0.00	0.00	0.00
+1 %	1.01000	2.5936	1.6917	-1.00	0.02	1.02

Table 3.4: Residual density descriptors for a change of the scale factor in simulated data on $\text{S}(\text{N}^t\text{Bu})_3$.

Both a decrease and an increase in the scale factor lead to a decrease in $d^f(0)$, while the other descriptors, e_{gross} and $\Delta\rho_0$, both increase with a changed (decreased as well as increased) scale factor. This behavior indicates an increase of features in the residual density. Thus, the optimum is characterized by extremal values of the descriptors.

Manipulation of anisotropic displacement parameters

The atoms in a molecule are not at fixed locations in the crystal as they are vibrating around their equilibrium positions. This leads to a statistically more diffuse distribution of the electrons with respect to the mean nuclear positions. Thus, the scattering power of the electrons decreases faster with increasing Bragg angle. This decrease is expressed by an exponential function. If the vibration of the atoms is

assumed to be isotropic the following equation is valid

$$f_{iso} = f e^{-2\pi^2 U d^{*2}} \quad (3.17)$$

with f_{iso} being the corrected atomic scattering factor, f being the original atomic scattering factor, U being the isotropic displacement parameter and d^* being the reciprocal of the resolution d (the distance between two lattice planes). To describe the motion of the atoms more realistically it is necessary to assume an anisotropic displacement

$$f_{aniso} = f e^{-2\pi^2 (U_{11} h^2 a^{*2} + U_{22} k^2 b^{*2} + U_{33} l^2 c^{*2} + 2U_{23} k l b^* c^* + 2U_{13} h l a^* c^* + 2U_{12} h k a^* b^*)} \quad (3.18)$$

where f_{aniso} is the corrected atomic scattering factor, the U_{ij} denote the anisotropic displacement parameters, h , k and l are the Miller indices, and a^* , b^* and c^* are the reciprocals of the cell axes a , b and c .

As can be seen from Eqs. 3.17 and 3.18 a larger thermal displacement parameter leads to a reduction of the atomic form factor for $d^* > 0$ and thus to a lower intensity. So, artificially increasing the U_{ij} decreases the F_{calc} . This effect is even stronger at higher scattering angles as the motion approaches the atoms more to each other and virtually increases the resolution which leads to an additional phase shift. Also, by an increase of the thermal displacement parameter the density is shifted further away from the nucleus which decreases F_{calc} and leads to a positive residual density near the atom and to a slightly (less distinct) negative residual density in regions further away from it.

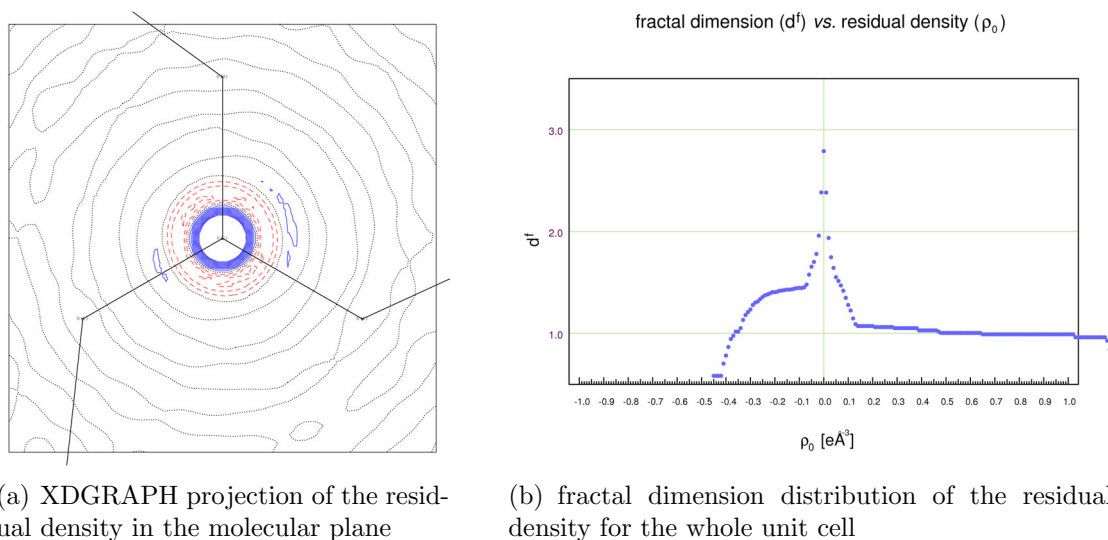


Figure 3.24: Residual density after an increase in the U_{ij} of S in $S(N^tBu)_3$ by 5%; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

	U_{11} (\AA^2)	U_{22} (\AA^2)	U_{33} (\AA^2)	U_{12} (\AA^2)	U_{13} (\AA^2)	U_{23} (\AA^2)
-10 %	0.010180	0.011129	0.021618	-0.005069	0.000121	-0.006068
-5 %	0.010745	0.011748	0.022819	-0.005350	0.000127	-0.006405
-2 %	0.011085	0.012119	0.023540	-0.005519	0.000131	-0.006607
-1 %	0.011198	0.012242	0.023780	-0.005576	0.000133	-0.006675
orig	0.011311	0.012366	0.024020	-0.005632	0.000134	-0.006742
+1 %	0.011424	0.012490	0.024260	-0.005688	0.000135	-0.006809
+2 %	0.011537	0.012613	0.024500	-0.005745	0.000137	-0.006877
+5 %	0.011877	0.012984	0.025221	-0.005914	0.000141	-0.007079
+10 %	0.012442	0.013603	0.026422	-0.006195	0.000147	-0.007416

Table 3.5: Change of U_{ij} of S in $S(N^tBu)_3$.

	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ (eÅ ⁻³)	$\rho_{0,max}$ (eÅ ⁻³)	$\Delta\rho_0$ (eÅ ⁻³)
-10 %	2.7891	4.3730	-6.04	1.00	7.04
-5 %	2.7894	2.1407	-2.96	0.49	3.45
-2 %	2.7930	0.8759	-1.17	0.19	1.36
-1 %	2.7962	0.5208	-0.59	0.10	0.69
orig	2.7972	0.3136	-0.00	0.00	0.00
+1 %	2.7954	0.5273	-0.09	0.58	0.67
+2 %	2.7930	0.8754	-0.19	1.15	1.34
+5 %	2.7880	2.1070	-0.45	2.84	3.29
+10 %	2.7828	4.5341	-0.87	5.49	6.36

Table 3.6: Residual density descriptors for a change of U_{ij} of S in simulated data of S(N^tBu)₃.

Fig. 3.24 shows the residual density in the main plane of the molecule (Fig. 3.24(a)) and the fractal dimension distribution for the whole unit cell (Fig. 3.24(b)) for an artificial increase in the anisotropic displacement parameters by 5 %. Fig. 3.24(a) shows a high concentration of positive residual density at the center of the atom, surrounded by a large amount of negative residual density, which again is surrounded by positive residual density. This behavior repeats by moving away further from the center of the atom.

A decrease in the U_{ij} would have the opposite effect, namely a decrease in the residual density near the core followed by an increase and so on.

When the anisotropic displacement parameters are changed $d^f(0)$ decreases while the other residual density descriptors are increased. This behavior reflects the increase in the features. Again, the optimum value is characterized by extremal values of the descriptors.

κ effects

κ is a radial screening parameter which considers the expansion or contraction of the spherical valence density P_ν in the multipole refinement. A value less than 1 denotes an expansion of the density whereas a value greater than 1 models a contraction of the spherical valence density.

$$\rho(\mathbf{r}) = \rho_c(r) + P_\nu \rho_\nu(\kappa r) + \rho_d(\kappa' \mathbf{r}) \quad (3.19)$$

The value for κ of the sulfur atom in S(N^tBu)₃ was changed to different values

(increased as well as decreased) and held fixed at this value (the values are shown in Table 3.7).

The plot for the residual density in the main plane of the molecule for a decrease in κ by 20 % is shown in Fig. 3.25(a) and the fractal dimension distribution of the corresponding residual density in the unit cell is shown in Fig. 3.25(b).

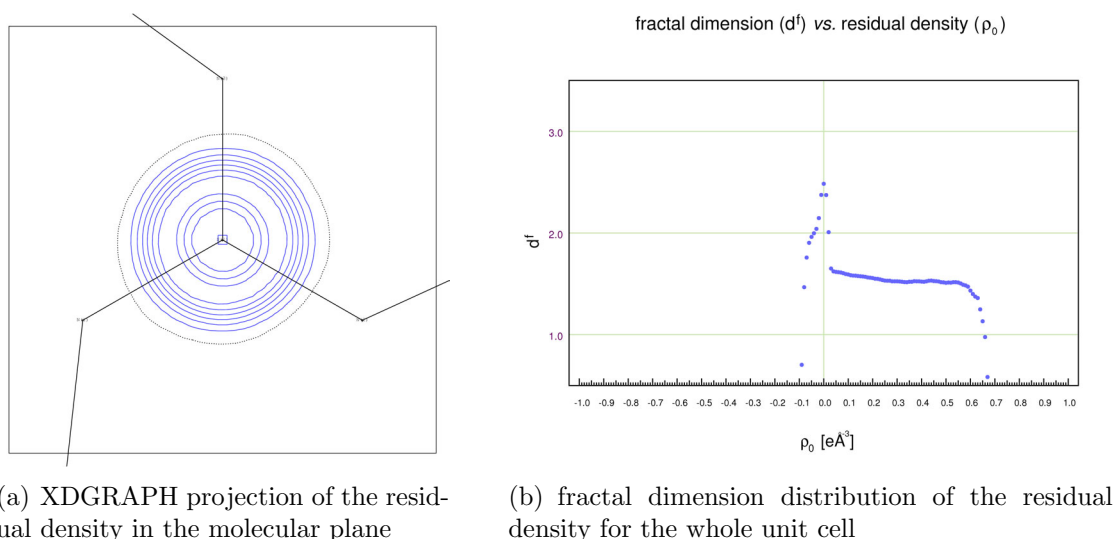


Figure 3.25: Residual density after a decrease in the κ parameter of S in $S(N^tBu)_3$ by 20 %; (a) blue solid lines: positive residual density, gray dotted line: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

	κ	$d^f(0)$	e_{gross}	$\rho_{0,min}$	$\rho_{0,max}$	$\Delta\rho_0$
			(e)	(eÅ ⁻³)	(eÅ ⁻³)	(eÅ ⁻³)
-50 %	0.544103	2.4199	13.7135	-0.15	1.31	1.46
-20 %	0.870566	2.4840	4.9861	-0.09	0.67	0.76
-10 %	0.979386	2.5467	2.3248	-0.04	0.37	0.41
-1 %	1.077325	2.7762	0.4098	-0.01	0.04	0.05
orig	1.088207	2.7972	0.3136	-0.00	0.00	0.00
+1 %	1.099089	2.7773	0.4091	-0.04	0.01	0.05
+10 %	1.197028	2.6117	1.8206	-0.44	0.05	0.49
+20 %	1.305848	2.5418	3.7344	-0.96	0.10	1.06
+50 %	1.632311	2.5330	10.0942	-2.89	0.27	3.16

Table 3.7: Residual density descriptors for a change of κ of S in simulated data of $S(N^tBu)_3$.

When the radial screening parameter κ at the sulfur atom is decreased the density is distributed in space more extended and diffuse. As a consequence, the concentration of the density near the position of the nucleus is decreased and the concentration apart from the center of the atom is increased. This causes a high positive residual

density ($F_{obs} - F_{calc} > 0$) around the nucleus of the sulfur atom (blue solid lines in Fig. 3.25(a)). The shape of the fractal dimension distribution for a wrong value of κ turned out to be very characteristic.

It can be seen from Table 3.7 that an increase as well as a decrease in κ leads to a decrease in $d^f(0)$ and an increase in e_{gross} and $\Delta\rho_0$. This reflects the deviation of κ from the original ideal value.

κ' effects

κ' is the radial expansion/contraction parameter for the deformation valence density.

$$\rho(\mathbf{r}) = \rho_c(r) + P_\nu \rho_\nu(\kappa r) + \rho_d(\kappa' \mathbf{r}) \quad (3.20)$$

$$\rho_d(\kappa' \mathbf{r}) = \sum_l R_l(\kappa' r) \sum_{m=-l}^l P_{lm} y_{lm}\left(\frac{\mathbf{r}}{r}\right) \quad (3.21)$$

Just like the radial expansion/contraction parameter κ a value greater than 1 for κ' denotes a contraction of the deformation density whereas a value less than 1 denotes an expansion of the density. In principle each of the multipole parameters P_{lm} can be scaled with its own κ' parameter but in the case of this study only one κ' was used for all multipoles (l values from 0 to 4). The value was varied from -50% to $+50\%$ of the original value to investigate the effects originating from a wrong value of κ' on the residual density. The results are listed in Table 3.8 and the corresponding plots for an increase in κ' of 50% are shown in Fig. 3.26.

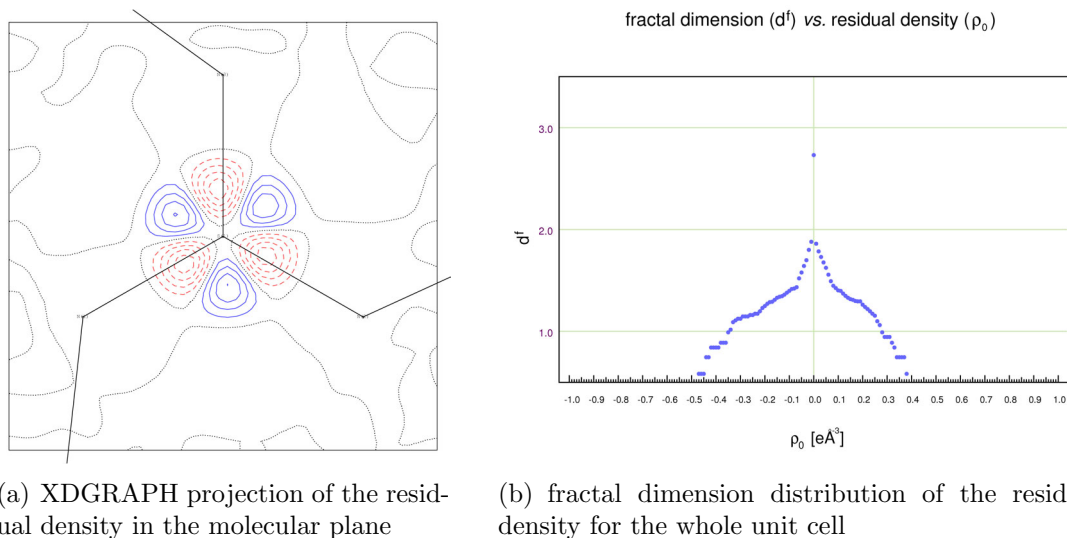


Figure 3.26: Residual density after an increase in κ' of S in $S(N^tBu)_3$ by 50%; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

	κ'	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($\text{e}\text{\AA}^{-3}$)	$\rho_{0,max}$ ($\text{e}\text{\AA}^{-3}$)	$\Delta\rho_0$ ($\text{e}\text{\AA}^{-3}$)
-50 %	0.596067	2.6386	1.3381	-0.20	0.24	0.44
-20 %	0.953708	2.7514	0.6623	-0.11	0.13	0.24
-10 %	1.072922	2.7637	0.5286	-0.04	0.12	0.16
-5 %	1.132528	2.7866	0.3743	-0.03	0.04	0.07
orig	1.192135	2.7972	0.3136	-0.00	0.00	0.00
+5 %	1.251742	2.7880	0.3709	-0.04	0.04	0.08
+10 %	1.311349	2.7783	0.4483	-0.08	0.07	0.15
+20 %	1.430562	2.7610	0.6115	-0.17	0.14	0.31
+50 %	1.788202	2.7303	1.1483	-0.48	0.39	0.87

Table 3.8: Residual density descriptors for a change of κ' of S in simulated data of $S(N^tBu)_3$.

As the deformation valence density is anisotropically distributed a deviation from the ideal value simultaneously increases the positive as well as the negative residual density. Density is shifted from regions where it is actually present to regions where less density should be present. An increase in κ' of the sulfur atom decreases the expansion of the deformation valence density corresponding to the multipoles which describe the threefold symmetry around the sulfur atom. As these multipoles shift density from the interbond regions to the bonding regions, an artificially increased κ' leads to a negative residual density along the bonds and a positive residual density between the bonds.

The threefold symmetry in the residual density is caused by constraints that were applied to the multipoles due to the threefold symmetry of the molecule and for the sake of reduction of parameters.

As was expected for the change of κ' towards values deviating from the ideal ones $d^f(0)$ decreases whereas e_{gross} and $\Delta\rho_0$ increase. The features in the residual density have increased as the changed model does not describe the data properly, anymore. In summary, the optimum is again characterized by extremal values of the descriptors.

Anomalous dispersion effects

When the energy of the photons in the X-ray beam is in resonance with the continuum excitation energy level of an electron in the crystal this leads to ionization by absorption of a photon *e.g.* by removing an electron from the K-edge of an atom, and subsequent emission of X-rays, *e.g.* $K\alpha$ radiation, by electronic relaxation. Due to the implied energy transfer, this is an inelastic scattering process called anomalous dispersion as opposed to the elastic scattering which does not change the energy of the photons. Anomalous dispersion causes an extra phase shift and a change in intensity. It consists of a real part $\Delta f'$ which is parallel to the amplitude and an imaginary part $\Delta f''$ which is perpendicular to the amplitude. The real part adds a positive or (which occurs more often) negative value to the atomic scattering factor f whereas the imaginary part always causes a positive phase shift.

$$f_{anom} = f + \Delta f' + \Delta f'' \quad (3.22)$$

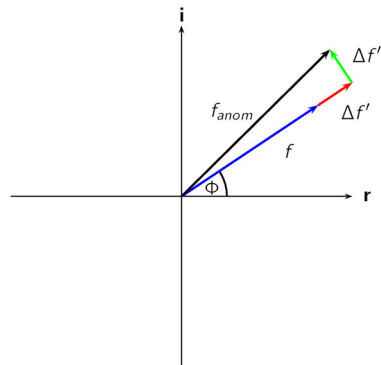


Figure 3.27: Change of the atom form factor when anomalous dispersion occurs.

For a certain atom and a given wavelength of the X-rays these values are constant and are tabulated *e.g.* in Table 4.2.6.8 in the International Tables Volume C [51]. The anomalous scattering of sulfur at Mo-K α radiation is slightly positive for both $\Delta f'$ and $\Delta f''$ and as the values are relatively small but significant a change should have a considerable impact on the residual density.

The $\Delta f'$ and $\Delta f''$ values were changed in the SCAT table in the xd.mas file (DELf' and DELf'', respectively) and a calculation of the structure factors was performed (XDLSM with the option "cycle 0").

Both contributions to the anomalous scattering $\Delta f'$ and $\Delta f''$ were varied simultaneously and independent of each other to investigate the individual influence of the parameters on the residual density.

Simultaneous change of $\Delta f'$ and $\Delta f''$ For the simultaneous variation of $\Delta f'$ and $\Delta f''$ both values were changed in the SCAT table in the xd.mas file. The resulting values are shown in Table 3.9 and Fig. 3.28 shows the residual density in the main plane of the molecule and the fractal dimension distribution for the whole unit cell, respectively.

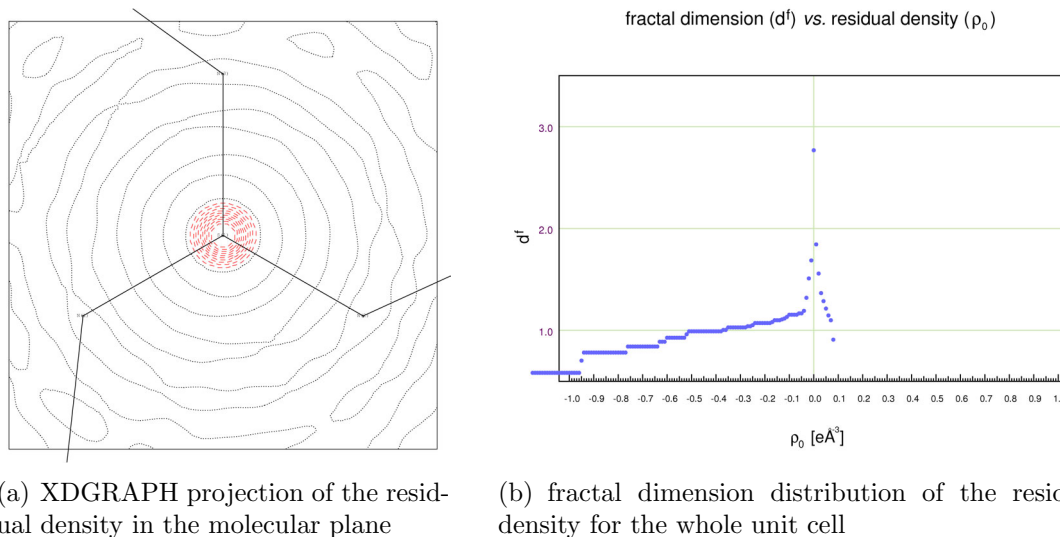


Figure 3.28: Residual density after an increase in $\Delta f'$ and $\Delta f''$ of S in S(N^tBu)₃ by 50 %; (a) red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 eÅ⁻³.

	$\Delta f'/\Delta f''$	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ (eÅ ⁻³)	$\rho_{0,max}$ (eÅ ⁻³)	$\Delta\rho_0$ (eÅ ⁻³)
-50 %	0.0623/0.0617	2.7628	0.8652	-0.09	1.16	1.25
-20 %	0.0997/0.0987	2.7832	0.4593	-0.04	0.46	0.50
-10 %	0.1121/0.1111	2.7921	0.3633	-0.02	0.23	0.25
orig	0.1246/0.1234	2.7972	0.3136	-0.00	0.00	0.00
+10 %	0.1371/0.1357	2.7937	0.3645	-0.23	0.02	0.25
+20 %	0.1495/0.1481	2.7854	0.4606	-0.46	0.04	0.50
+50 %	0.1869/0.1851	2.7675	0.8785	-1.16	0.09	1.25

Table 3.9: Residual density descriptors for a change of $\Delta f'$ and $\Delta f''$ of S in simulated data of S(N^tBu)₃.

An increase in $\Delta f'$ and $\Delta f''$ of S increases the atomic structure factor of the sulfur atoms. This leads to an increase of F_{calc} and thus an increase in the negative residual density. This can be seen from Fig. 3.28 as negative residual density appears near the nucleus of the sulfur atom. The residual density descriptors also reflect the wrong description of the data. As can be seen (Table 3.9), both an increase and a decrease in $\Delta f'$ and $\Delta f''$ decrease the fractal dimension $d^f(0)$, whereas e_{gross} and $\Delta\rho_0$ are increased.

Change of $\Delta f'$ To investigate the real and the imaginary part of the anomalous dispersion separately both values were varied independent of each other.

For investigating the pure influence of the real part only $\Delta f'$ was varied in the SFAC section of the xd.mas file while $\Delta f''$ was kept at its original value.

Fig. 3.29(a) shows the residual density distribution in the main plane of the molecule for an increase in $\Delta f'$ about 50 % and Fig. 3.29(b) shows the respective fractal dimension distribution for the whole unit cell. In Table 3.10 the results of variations in $\Delta f'$ from -50 % to +50 % are given.

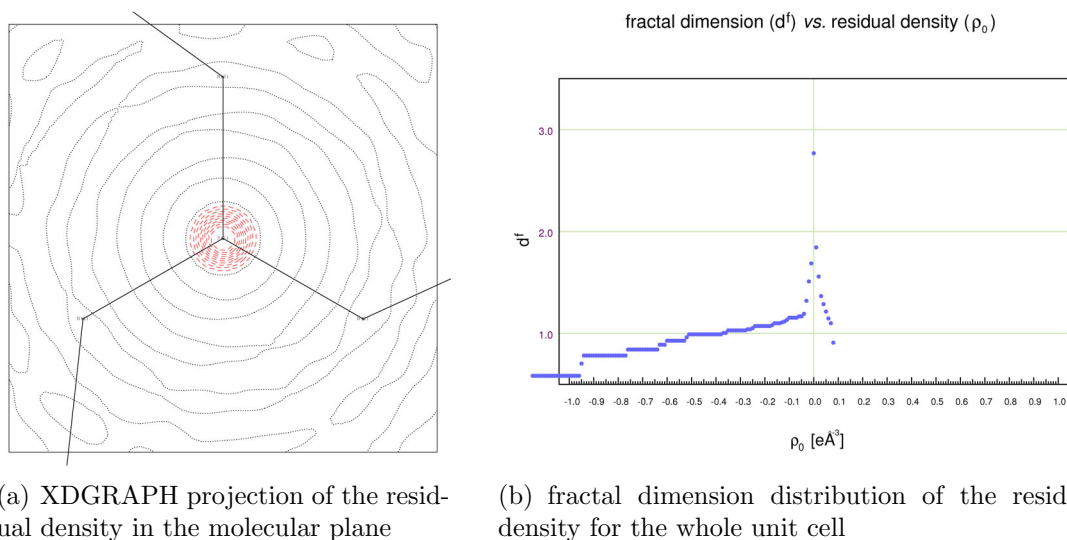


Figure 3.29: Residual density after an increase in $\Delta f'$ of S in $\text{S}(\text{N}^t\text{Bu})_3$ by 50 %; (a) red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

	$\Delta f' / \Delta f''$	$d^f(0)$	e_{gross}	$\rho_{0,min}$	$\rho_{0,max}$	$\Delta \rho_0$
			(e)	($\text{e}\text{\AA}^{-3}$)	($\text{e}\text{\AA}^{-3}$)	($\text{e}\text{\AA}^{-3}$)
-50 %	0.0623/0.1234	2.7735	0.7480	-0.09	1.13	1.22
-20 %	0.0997/0.1234	2.7881	0.4304	-0.04	0.45	0.49
-10 %	0.1121/0.1234	2.7934	0.3553	-0.02	0.23	0.25
orig	0.1246/0.1234	2.7972	0.3136	-0.00	0.00	0.00
+10 %	0.1371/0.1234	2.7956	0.3562	-0.23	0.02	0.25
+20 %	0.1495/0.1234	2.7894	0.4334	-0.45	0.03	0.48
+50 %	0.1869/0.1234	2.7675	0.8785	-1.16	0.09	1.25

Table 3.10: Residual density descriptors for a change of $\Delta f'$ of S in simulated data of $\text{S}(\text{N}^t\text{Bu})_3$.

Fig. 3.28 is reminiscent of Fig. 3.29 and so are Tables 3.9 and 3.10. This suggests that $\Delta f'$ is the dominating factor.

Change of $\Delta f''$ For the sole variation of $\Delta f''$ the corresponding entry in the SFAC section in the xd.mas file was varied whereas $\Delta f'$ was kept fixed. Fig. 3.30(a) shows the residual density plot for the main plane of the molecule and Fig. 3.30(b) shows the respective fractal dimension distribution. In Table 3.11 all values for the variation of $\Delta f''$ from -50 % to +50 % are listed.

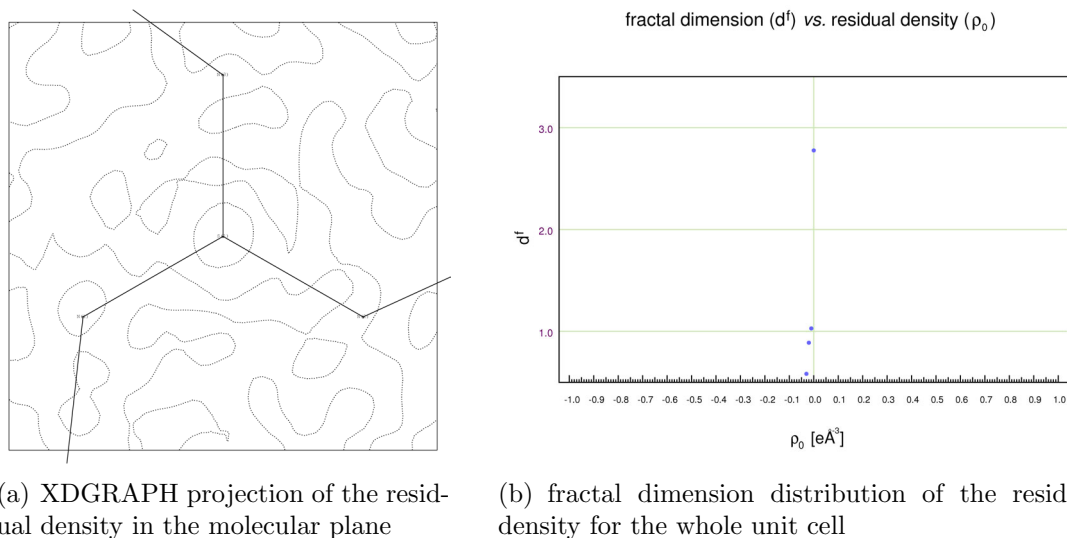


Figure 3.30: Residual density after an increase in $\Delta f''$ of S in $S(N^tBu)_3$ by 50 %; (a) gray dotted lines: zero residual density.

	$\Delta f' / \Delta f''$	$d^f(0)$	e_{gross}	$\rho_{0,min}$	$\rho_{0,max}$	$\Delta\rho_0$
			(e)	($e\text{\AA}^{-3}$)	($e\text{\AA}^{-3}$)	($e\text{\AA}^{-3}$)
-50 %	0.1246/0.0617	2.7774	0.4047	-0.01	0.02	0.03
-20 %	0.1246/0.0987	2.7917	0.3324	-0.00	0.01	0.01
-10 %	0.1246/0.1111	2.7956	0.3186	-0.00	0.01	0.01
orig	0.1246/0.1234	2.7972	0.3136	-0.00	0.00	0.00
+10 %	0.1246/0.1357	2.7962	0.3194	-0.01	0.00	0.01
+20 %	0.1246/0.1481	2.7926	0.3359	-0.01	0.00	0.01
+50 %	0.1246/0.1851	2.7758	0.4392	-0.04	0.01	0.05

Table 3.11: Residual density descriptors for a change of $\Delta f''$ of S in simulated data of $S(N^tBu)_3$.

The plots in Fig. 3.30 and the values in Table 3.11 confirm the principal influence of the real part of the anomalous dispersion $\Delta f'$ on the residual density.

Although in general both parameters $\Delta f'$ and $\Delta f''$ affect both the amplitude and the phase of the structure factors, and both the amplitude and the phase contribute to the residual density, only the change of $\Delta f'$ had a significant influence on the residual density.

For simultaneous as well as independent changes in $\Delta f'$ and $\Delta f''$ it was observed that all residual density descriptors behave in an expected way, *i.e.* $d^f(0)$ decreases whereas e_{gross} and $\Delta\rho_0$ increase. They perfectly account for the deterioration in the parameters due to the manipulated values.

3.5.2 Application of new descriptors to experimental data

3.5.2.1 Progression of a multipole refinement

During a multipole refinement several parameters are introduced which describe the distribution of the electron density more adequately than a conventional Independent Atom Model (IAM). Normally, these parameters are added and refined in a stepwise manner because otherwise convergence might not be reached when some parameters are strongly correlated. The parameters to be refined are the coordinates and thermal motion parameters as well as the radial extension/contraction parameters κ and κ' for the spherical valence density and the deformation density and the monopole- and multipole population parameters P_ν and P_{lm} , respectively. To investigate the influence of the individual parameters on the residual density it was analyzed after each step of the multipole refinement of $S(N^tBu)_3$ [22] (see Fig. 3.31). Data of the fully refined IAM were used and the individual instruction files (xd.mas) for the multipole refinement were kindly provided by Dr. Dirk Leusser.

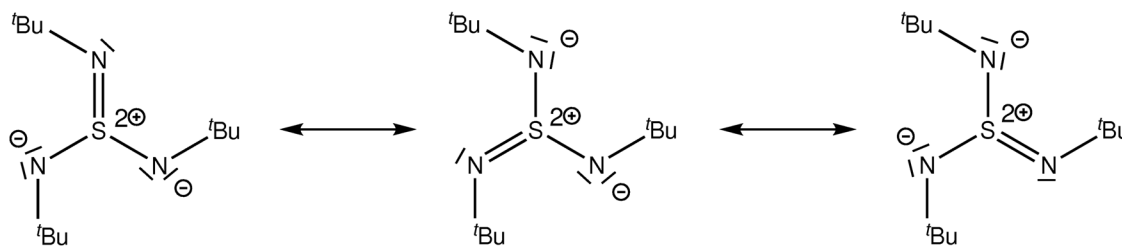


Figure 3.31: Lewis diagrams of the most dominant mesomeric structures of $S(N^tBu)_3$.

For each refinement step a least-squares refinement against F^2 with XDLSM was performed with ten cycles or until convergence was reached. Only data with $I > 3\sigma(I)$ were used for the refinement. The residual density grid was calculated with XDFOUR for a $50 \times 50 \times 50$ grid over the whole unit cell and no resolution cutoff was applied for the Fourier transformation. The fractal dimension $d^f(0)$ for the starting model was 2.6681, e_{gross} was 11.5432 e and the flatness $\Delta\rho_0$ was $1.29 \text{ e}\text{\AA}^{-3}$. Eight main refinement steps were applied. The results of the residual density analysis are illustrated in Fig. 3.32 and listed in Table 3.12.

In the first step the scale factor as well as the coordinates of all atoms, the anisotropic displacement parameters for the non-hydrogen (non-H) atoms and the isotropic displacement parameters for the H atoms were refined for the IAM model from the SHELX refinement with conventional scattering factors (IAM). From the fractal dimension distribution it can be seen that there are still many errors which influence the negative as well as the positive residual density. In the IAM the atoms

are considered uncharged and spherical, therefore, neither charge transfer nor re-distribution of electrons due to chemical bonding are accounted for. This causes shoulders in the positive residual density region ($F_{obs} > F_{calc}$ due to bonding and non-bonded electrons which are not described by the model) as well as in the negative residual density region ($F_{obs} < F_{calc}$ because all electrons around the atoms were placed close to the nuclei).

In the next step a high order ($\sin \theta/\lambda > 0.6 \text{ \AA}^{-1}$) refinement of the scale factor, the coordinates and the vibrational parameters of the non-hydrogen atoms was performed ($xyz + U_{ij}$ non-H ($\sin \theta/\lambda > 0.6 \text{ \AA}^{-1}$)). As the information about the inner electrons is contained mainly in the high resolution reflections^[52] the coordinates can be determined more exactly when the resolution is restricted to high order data (high-pass filtering). On the other hand this also means that less information about the bonding electrons is contained in the high order data which is the reason that no features in the residual density distribution are seen and $d^f(0)$ increased to 2.7669, e_{gross} decreased to 4.9590 e and $\Delta\rho_0$ decreased to 0.45 e\AA^{-3} . As will be seen in section 3.5.2.3 a truncation in the high order data leads to an improvement in the residual density descriptors for two independent reasons: the contribution to the error resulting from the difference $F_{obs} - F_{calc}$ is effectively zero for an omitted observation. This holds not only true for the truncation of the high order data but also for the low order data as was the case here. Furthermore, when the number of model parameters is kept fixed and the number of observations is reduced, the flexibility of the model has effectively increased. This leads to an additional improvement if the model parameters are adjusted. Both effects improve the agreement between model and data. There is, however, one important difference in the truncation of high order data and low order data: the first acts like a low-pass filter, *i.e.* only the low frequency components pass and the high frequency components are omitted. This leads to a smoother distribution of the residual density, averaging out the details and emphasizing the main features. As this decreases the number of zero-crossings of the residual density the fractal dimension $d^f(0)$ is also decreased. For the high-pass filter, *i.e.* when the low order data are omitted, the opposite is the case: the main features in the residual density are omitted whereas the details are emphasized. This leads to an increase of zero-crossings of the residual density and $d^f(0)$ increases. Therefore, $d^f(0)$ increases for high-pass filtered data and decreases for low-pass filtered data, whereas e_{gross} and $\Delta\rho_0$ decrease only under data truncation. This can be seen from the comparison of row 2 in Table 3.12 and Table 3.22.

In a subsequent refinement step the scale factor, the monopole populations P_ν for all atoms and the expansion/contraction parameters κ for the spherical valence

density of the non-hydrogen atoms were refined against all data with $I > 3\sigma(I)$ for the whole resolution range ($P_\nu + \kappa$ non-H). The refinement leads to a pronounced reduction of the monopole populations for few atoms (the sulfur atom and the carbon atoms) and to a slight increase for many atoms (nitrogen and hydrogen atoms). This step reduces mainly the negative residual density as the monopoles and κ only account for the amount and the radial distribution of the valence electrons but not for their angular distribution. Thus, the polarization of the pseudo atoms is still not described properly, however, charge transfer and a change in the radial functions is accounted for. Compared to the first refinement step, the fractal dimension decreases and the gross residual electrons increase, so both indicate a deterioration of the model ($d^f(0) = 2.6111$, $e_{gross} = 13.4689$ e), whereas the flatness increases ($\Delta\rho_0 = 0.85$ eÅ⁻³) and so indicates an improvement of the model. These apparently contradicting results reflect both the model improvement (charge transfer and the radial functions) and deficiency (the lack of allowing for non-spherical changes).

In consequence, taking asphericities into account by introducing multipole population parameters P_{lm} and a refinement together with the scale factor and the monopole population parameters P_ν for all atoms, the distribution of the residual density shows a clear improvement of the model. There are no distinct shoulders in the positive nor in the negative residual density region anymore ($P_\nu + P_{lm}$ all). The residual density descriptors also indicate a progress in the model ($d^f(0) = 2.6592$, $e_{gross} = 9.0711$ e and $\Delta\rho_0 = 0.71$ eÅ⁻³). The distribution of the electron density in the bonds and of the non-bonding electrons is now modeled much more adequately.

The refinement of the scale factor and the expansion/contraction parameter κ' of the aspherical deformation density of the non-hydrogen atoms (κ' non-H) does not result in a great improvement of the model and the residual density descriptors ($d^f(0) = 2.6596$, $e_{gross} = 8.9324$ e and $\Delta\rho_0 = 0.71$ eÅ⁻³).

In the next step the scale factor and the coordinates of all atoms as well as the anisotropic displacement parameters for the non-hydrogen atoms and the isotropic displacement parameters for hydrogen atoms were refined. The hydrogen atoms were shifted along the C–H vectors to distances derived from neutron diffraction experiments (C–H distance of 1.085 Å^[53]) ($xyz + U_{ij}$ non-H + RESET). The flatness increases to $\Delta\rho_0 = 0.67$ eÅ⁻³, but as the description of the monopoles and multipoles of the hydrogen atoms does not fit to the changed coordinates anymore, a shoulder in the positive residual density region appears again and a deterioration of the model is indicated by the decrease in the fractal dimension ($d^f(0) = 2.6133$) and the increase in the gross residual electrons ($e_{gross} = 12.1039$ e).

When the scale factor, the monopole- and the multipole population parameters for all atoms and the expansion/contraction parameters κ for the non-hydrogen atoms are refined ($P_\nu + P_{lm} + \kappa$), the model is given the possibility to account for charge transfer and asphericities around all atoms simultaneously, resulting in a distinct reduction of the residual density. The fractal dimension increases to $d^f(0) = 2.6918$, the gross residual electrons decrease to $e_{gross} = 7.3794\text{e}$ and the flatness is increased to $\Delta\rho_0 = 0.61\text{e}\text{\AA}^{-3}$.

The final refinement of the scale factor, P_ν , P_{lm} , the expansion/contraction parameters for the spherical valence (κ) and the aspherical deformation density (κ') of the non-hydrogen atoms ($P_\nu + P_{lm} + \kappa + \kappa'$) leads to the highest value for the fractal dimension ($d^f(0) = 2.6922$), the lowest value for the gross residual electrons ($e_{gross} = 7.1992\text{e}$) and the best value for the flatness with $\Delta\rho_0 = 0.57\text{e}\text{\AA}^{-3}$ (apart from the second step in which Fourier truncation is involved).

In total, the fractal dimension increased from 2.6681 to 2.6922, the gross residual electrons decreased from 11.5432e to 7.1992e and the flatness increased from $1.29\text{e}\text{\AA}^{-3}$ to $0.57\text{e}\text{\AA}^{-3}$. The improvement in the residual density can also be seen in Fig. 3.33 which represents the residual density in a plane containing the heaviest atoms in the molecule before (Fig. 3.33(a)) and after (Fig. 3.33(b)) the multipole refinement. This clearly shows that the multipole model describes the electron density more adequate than the spherical Independent Atom Model. However, the shape of the fractal dimension distribution is not parabolic, which should be the case if the remaining errors came from Gaussian noise in the data, only. This deviation in the shape from the ideal case results from a disorder in the molecule by rotation about 60° around the sulfur atom which was too small to be modeled but big enough to show up in the residual density (the occupation factor of the second position is assumed to be less than 1%).

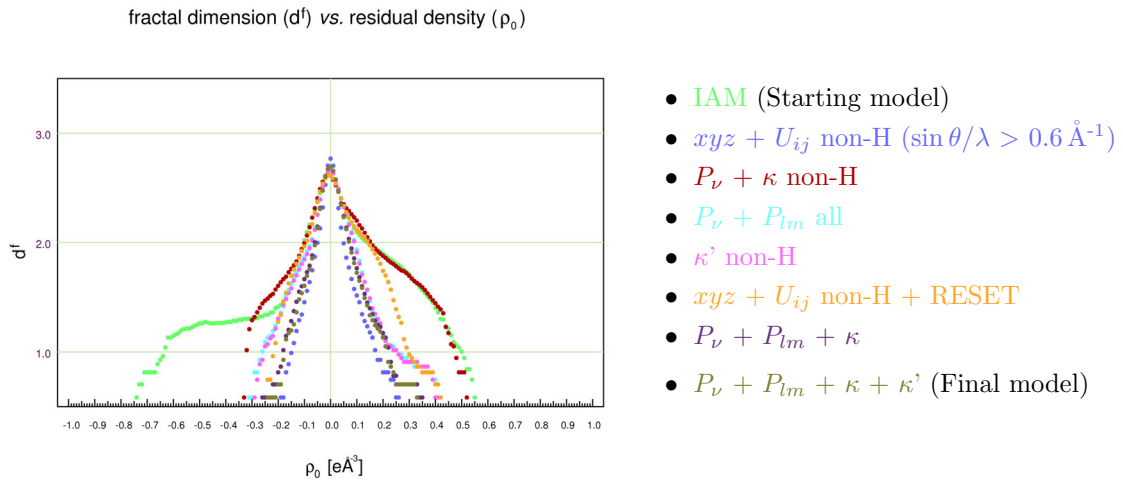


Figure 3.32: Fractal dimension distribution for the individual steps of a multipole refinement on $S(N^tBu)_3$.

	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
IAM	2.6681	11.5432	-0.74	0.55	1.29
$xyz + U_{ij}$ non-H ($\sin \theta/\lambda > 0.6 \text{ \AA}^{-1}$)	2.7669	4.9590	-0.19	0.26	0.45
$P_\nu + \kappa$ non-H	2.6111	13.4689	-0.33	0.52	0.85
$P_\nu + P_{lm}$ all	2.6592	9.0711	-0.31	0.40	0.71
κ' non-H	2.6596	8.9324	-0.29	0.42	0.71
$xyz + U_{ij}$ non-H + RESET	2.6133	12.1039	-0.25	0.42	0.67
$P_\nu + \kappa + P_{lm}$	2.6918	7.3794	-0.26	0.35	0.61
$P_\nu + P_{lm} + \kappa + \kappa'$	2.6922	7.1992	-0.24	0.33	0.57

Table 3.12: Residual density descriptors applied to the individual steps of a multipole refinement on $S(N^tBu)_3$.

Fig. 3.33 shows the residual densities around the sulfur atom in $S(N^tBu)_3$ before the IAM refinement (Fig. 3.33(a)) and after the multipole refinement (Fig. 3.33(b)). The green and the olive green lines in Fig. 3.32 show the fractal dimension distribution for the residual densities for the whole unit cell corresponding to the residual densities in Figs. 3.33(a) and 3.33(b), respectively.

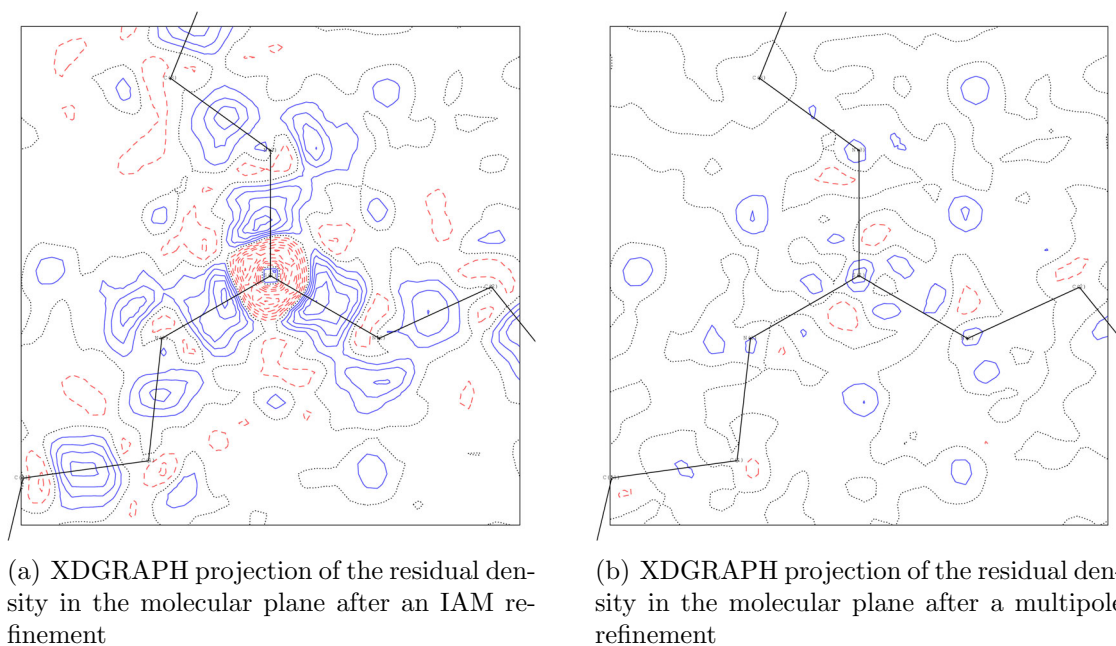


Figure 3.33: Residual density before and after the multipole refinement for $S(N^tBu)_3$; blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing $0.1 \text{ e}\text{\AA}^{-3}$.

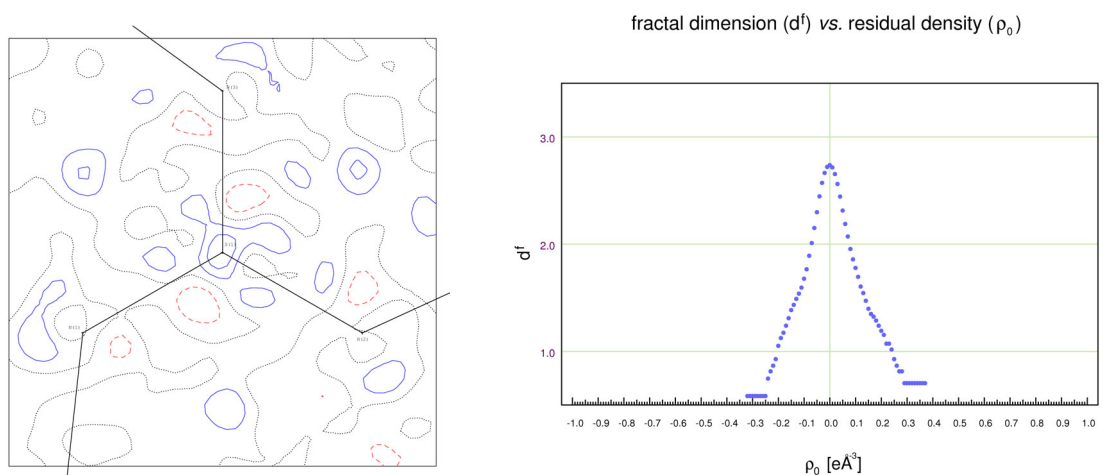
3.5.2.2 Analysis of the impact of a change of model parameters on the residual density

For a comparison of the effect of the change of multipole parameters on ideal simulated and on experimental data, the same parameters as in section 3.5.1.3 were changed for the fully refined multipole model of $S(N^tBu)_3$. For the simulated data the only source of error was the respective manipulated parameter, thus the change in the residual density is caused only by this one error. For experimental data, however, not all the errors are known and thus it cannot be predicted how exactly the residual density is influenced due to a possible interference of several sources of error.

The respective parameters were again the scale factor k , the anisotropic displacement parameters U_{ij} of the sulfur atom, the expansion/contraction parameters of the spherical valence (κ) and the aspherical deformation density (κ') of the sulfur atom and the anomalous dispersion values $\Delta f'$ and $\Delta f''$ of the sulfur atom. The original values for the respective parameters are given in Table 3.13. The starting distribution of the residual density is shown in Fig. 3.34(a) for the main plane of the molecule and in Fig. 3.34(b) for a $50 \times 50 \times 50$ grid over the whole unit cell.

For all subsequent parameter manipulations the parameters were again all changed independent of each other and no refinement was performed (“cycle 0” in XDLSM). Two Fourier transformations with XDFOUR were applied to calculate one grid with

50×50 data points in a plane around the molecule and one grid of $50 \times 50 \times 50$ data points covering the complete unit cell. The figures on the left hand side show the residual density in the plane containing the molecule and were drawn with XDGRAPH. The figures on the right hand side show the fractal dimension distributions which were produced with jnk2RDA on the $50 \times 50 \times 50$ residual density grid for the whole unit cell.



(a) XDGRAPH projection of the residual density in the molecular plane

(b) fractal dimension distribution of the residual density for the whole unit cell

Figure 3.34: Residual density after the multipole refinement of $S(N^tBu)_3$; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

$d^f(0)$		2.7363		
e_{gross}	(e)	8.3961		
$\rho_{0,min}$	($\text{e}\text{\AA}^{-3}$)	-0.32		
$\rho_{0,max}$	($\text{e}\text{\AA}^{-3}$)	0.37		
$\Delta\rho_0$	($\text{e}\text{\AA}^{-3}$)	0.69		
κ		1.088409		
κ'		1.188396		
U_{ij}	(\AA^2)	U_{11} : 0.011310	U_{22} : 0.012365	U_{33} : 0.024019
		U_{12} : -0.005632	U_{13} : 0.000134	U_{23} : -0.006741
$\Delta f'$		0.1246		
$\Delta f''$		0.1234		
scale factor $k1$		0.188459E+01		
scale factor $k2$		0.209647E+01		

Table 3.13: Original values of S in $S(N^tBu)_3$; the grid resolution was $50 \times 50 \times 50$.

Manipulation of the scale factor k

As the data for $S(N^tBu)_3$ were collected at two different detector positions (-31° and -80° in 2θ , respectively), the refinement was performed in two different batches. The low order batch contains the data with $\sin\theta/\lambda < 0.625 \text{ \AA}^{-1}$ while the high order batch describes data with $\sin\theta/\lambda > 0.625 \text{ \AA}^{-1}$ [22]. This procedure results in two different scale factors (k_1 and k_2) which in this study were changed simultaneously. Fig. 3.35(a) shows the main plane of the molecule after an increase in the two scale factors of 1%, in Fig. 3.35(b) the corresponding fractal dimension distribution is shown. From both figures it can be seen that the negative residual density increased which occurs mainly at the atomic positions.

A decrease in the scale factor(s) results in an increase in the positive residual density as F_{obs} is scaled to higher values which leads to positive values for $F_{obs} - F_{calc}$. The values for the residual density descriptors for both an increase as well as a decrease of the scale factor are given in Table 3.14.

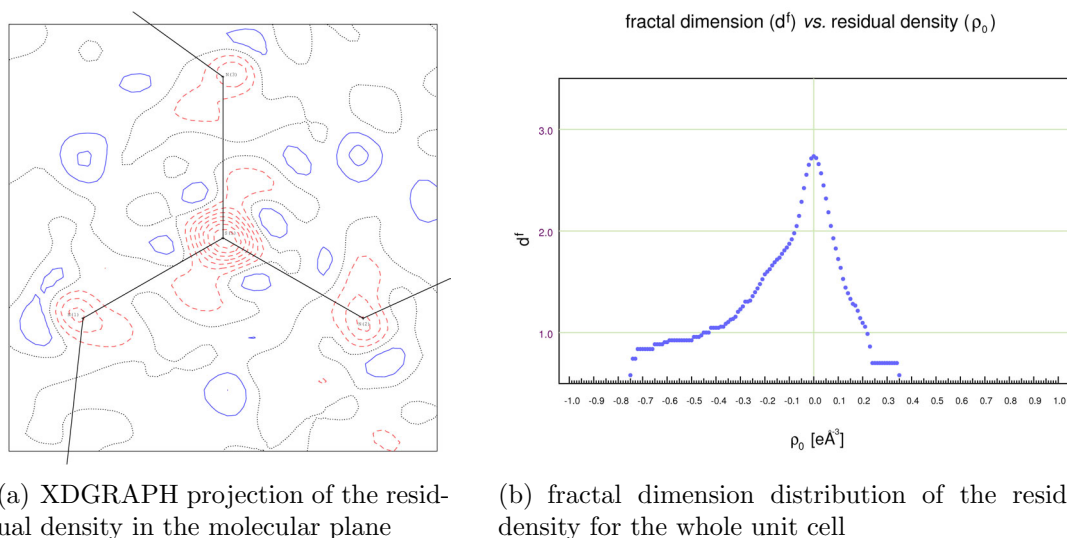


Figure 3.35: Residual density after an increase in the scale factors k_1 and k_2 in $S(N^tBu)_3$ by 1%; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 e\AA^{-3} .

	k_1/k_2	$d^f(0)$	e_{gross}	$\rho_{0,min}$	$\rho_{0,max}$	$\Delta\rho_0$
			(e)	(e\AA^{-3})	(e\AA^{-3})	(e\AA^{-3})
-1 %	1.86574/2.07551	2.7326	8.8613	-0.28	1.31	1.59
orig	1.88459/2.09647	2.7363	8.3961	-0.32	0.37	0.69
+1 %	1.90344/2.11743	2.7342	8.5977	-0.75	0.35	1.10

Table 3.14: Residual density descriptors for a change of both scale factors for experimental data of $S(N^tBu)_3$.

Both a decrease and an increase in the scale factor lead to a decrease in $d^f(0)$, while the other descriptors, e_{gross} and $\Delta\rho_0$, both increase with a changed (decreased as well as increased) scale factor. This behavior indicates an increase of features in the residual density. Just like in the case of simulated data the optimum is characterized by extremal values of the descriptors.

Manipulation of anisotropic displacement parameters

The anisotropic displacement parameters U_{ij} of the sulfur atom were changed in the range from -50% to $+50\%$ (see Table 3.15). Fig. 3.36 shows the residual density in the plane of the molecule (Fig. 3.36(a)) and the fractal dimension distribution for the whole unit cell (Fig. 3.36(b)) for an artificial increase in the anisotropic displacement parameters of 5%. Fig. 3.36(a) shows a high concentration of positive residual density at the center of the atom, surrounded by a large amount of negative residual density, which again is surrounded by positive residual density. This behavior corresponds exactly to the behavior of the simulated data for a change in U_{ij} (see section 3.5.1.3).

A decrease in the U_{ij} has the opposite effect on the residual density and the respective descriptors, as can be seen from Table 3.16.

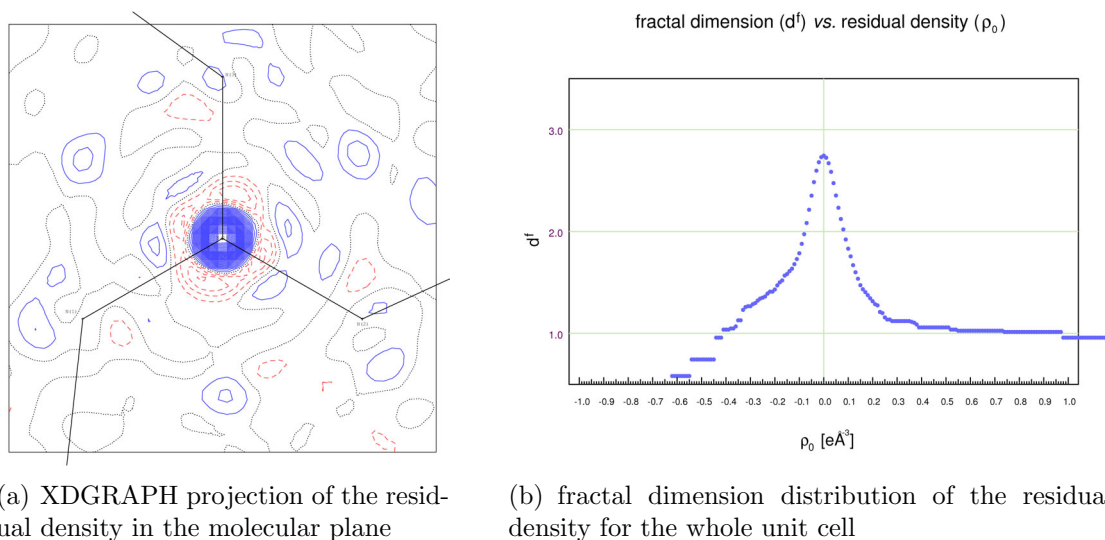


Figure 3.36: Residual density after an increase in the U_{ij} of S in $S(N^tBu)_3$ by 5%; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 eÅ^{-3} .

	U_{11} (\AA^2)	U_{22} (\AA^2)	U_{33} (\AA^2)	U_{12} (\AA^2)	U_{13} (\AA^2)	U_{23} (\AA^2)
-10 %	0.010179	0.011129	0.021617	-0.005069	0.000121	-0.006067
-5 %	0.010745	0.011747	0.022818	-0.005350	0.000127	-0.006404
-1 %	0.011197	0.012241	0.023779	-0.005576	0.000133	-0.006674
orig	0.011310	0.012365	0.024019	-0.005632	0.000134	-0.006741
+1 %	0.011423	0.012489	0.024259	-0.005688	0.000135	-0.006808
+5 %	0.011876	0.012983	0.025220	-0.005914	0.000141	-0.007078
+10 %	0.012441	0.013602	0.026421	-0.006195	0.000147	-0.007415

Table 3.15: Change of U_{ij} of S in $S(\text{N}^t\text{Bu})_3$.

	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($\text{e}\text{\AA}^{-3}$)	$\rho_{0,max}$ ($\text{e}\text{\AA}^{-3}$)	$\Delta\rho_0$ ($\text{e}\text{\AA}^{-3}$)
-10 %	2.7443	9.6512	-5.55	0.88	6.43
-5 %	2.7377	8.6671	-2.57	0.49	3.06
-1 %	2.7364	8.3773	-0.39	0.38	0.77
orig	2.7363	8.3961	-0.32	0.37	0.69
+1 %	2.7378	8.4636	-0.32	0.84	1.16
+5 %	2.7432	9.0837	-0.62	3.02	3.64
+10 %	2.7499	10.3525	-1.01	5.61	6.62

Table 3.16: Residual density descriptors for a change of U_{ij} of S in experimental data of $S(\text{N}^t\text{Bu})_3$.

When the anisotropic displacement parameters are changed all of the residual density descriptors are increased. For $d^f(0)$ it can be said that for small changes in the residual density (small changes in the U_{ij}) the introduction of new features increases the total amount of zero residual density which increases $d^f(0)$ although no improvement in the model was achieved. This demonstrates that for experimental data, where there is no full control over the errors that affect the residual density, not the absolute value of $d^f(0)$ is a measure of the features of the residual density, but the relative change to its optimum value: $d^f(0)$ assumes an extremal value for the optimal model, may it be a maximum or a minimum.

That $d^f(0)$ sometimes increases with a model of minor quality is in complete analogy to other measures of quality like the agreement factors, as sometimes a lower R -factor corresponds to a physically meaningless model. As an example, the displacement parameters may account for the disordered part of a structure. In this case it is in the responsibility of the crystallographer to reject the model despite its lower R -value. In much the same way, $d^f(0)$ just quantifies the total amount of

zero residual density regardless of its origin from a “natural” source (like noise) or from model artifacts. The source of the residual density, however, can be deduced from the shape of the residual density distribution. Furthermore, very often small changes in $d^f(0)$ are accompanied by large changes in e_{gross} and $\Delta\rho_0$.

The increase in e_{gross} and $\Delta\rho_0$ shows that the anisotropic displacement parameters had been refined to their ideal values. Obviously, an error of about 5% in U_{ij} is a massive model error, which would not arise naturally in a high resolution X-ray study.

κ effects

The value for κ of the sulfur atom in $S(N^tBu)_3$ was changed to different values (increased as well as decreased) and held fixed at this value (values shown in Table 3.17). A structure factor calculation with XDLSM with the option “cycle 0” was performed and a $50 \times 50 \times 50$ residual density grid for the whole unit cell was created with XDFOUR and was analyzed with jnk2RDA. The plot for the residual density in the main plane of the molecule for a decrease in κ by 20% is shown in Fig. 3.37(a) and the fractal dimension distribution of the corresponding residual density in the unit cell is shown in Fig. 3.37(b).

	κ	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
-50 %	0.544205	2.6223	16.8137	-0.30	1.56	1.86
-20 %	0.870727	2.7082	10.4451	-0.32	0.81	1.13
-10 %	0.979568	2.7273	9.0473	-0.33	0.55	0.88
-1 %	1.077525	2.7361	8.4040	-0.32	0.37	0.69
orig	1.088409	2.7363	8.3961	-0.32	0.37	0.69
+1 %	1.099293	2.7363	8.4065	-0.32	0.37	0.69
+10 %	1.197250	2.7320	8.9755	-0.60	0.37	0.97
+20 %	1.306091	2.7216	9.9671	-1.07	0.36	1.43
+50 %	1.632614	2.6779	13.9731	-2.87	0.37	3.24

Table 3.17: Residual density descriptors for a change of κ of S in experimental data of $S(N^tBu)_3$.

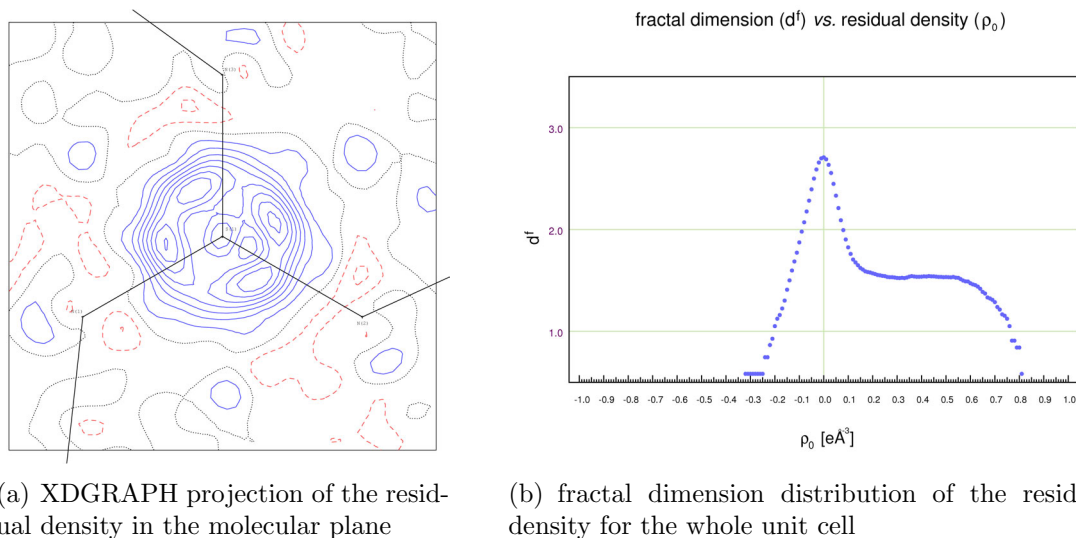


Figure 3.37: Residual density after a decrease in κ of S in $S(N^tBu)_3$ by 20%; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

The more extended and diffuse distribution of density causes a high positive residual density ($F_{obs} - F_{calc} > 0$) around the nucleus of the sulfur atom (blue solid lines in Fig. 3.37(b)) and a slightly negative residual density ($F_{obs} - F_{calc} < 0$) further away from it (red dashed lines in Fig. 3.37(b)). The shape of the fractal dimension distribution for a wrong value of κ turned out to be very characteristic.

In the case of the κ parameter it can be seen from Table 3.17 that an increase as well as a decrease in κ leads to a decrease in $d^f(0)$ and an increase in e_{gross} and $\Delta\rho_0$. This indicates that the original κ has already been at its optimal value.

κ' effects

Like in the study on simulated data only one value for κ' was used for all the multipoles (l going from 0 to 4). The value was varied from -50% to $+50\%$ of the original value to investigate the effects originating from a wrong value of κ' on the residual density. The results are listed in Table 3.18 and the corresponding plots for an increase in κ' of 50% are shown in Fig. 3.38 for a particular example.

	κ'	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
-50 %	0.594198	2.7339	8.6774	-0.33	0.38	0.71
-20 %	0.950717	2.7355	8.4761	-0.32	0.37	0.69
-10 %	1.069556	2.7361	8.4236	-0.32	0.37	0.69
-5 %	1.128976	2.7364	8.4057	-0.32	0.37	0.69
orig	1.188396	2.7363	8.3961	-0.32	0.37	0.69
+5 %	1.247816	2.7369	8.3966	-0.32	0.37	0.69
+10 %	1.307236	2.7373	8.4028	-0.32	0.37	0.69
+20 %	1.426075	2.7371	8.4381	-0.32	0.37	0.69
+50 %	1.782594	2.7348	8.6348	-0.46	0.49	0.95

Table 3.18: Residual density descriptors for a change of κ' of S in experimental data of $S(N^tBu)_3$.

Again, the threefold symmetry in the residual density plot (Fig. 3.38) reflects the threefold symmetry that was applied for constraining the multipole population parameters P_{lm} .

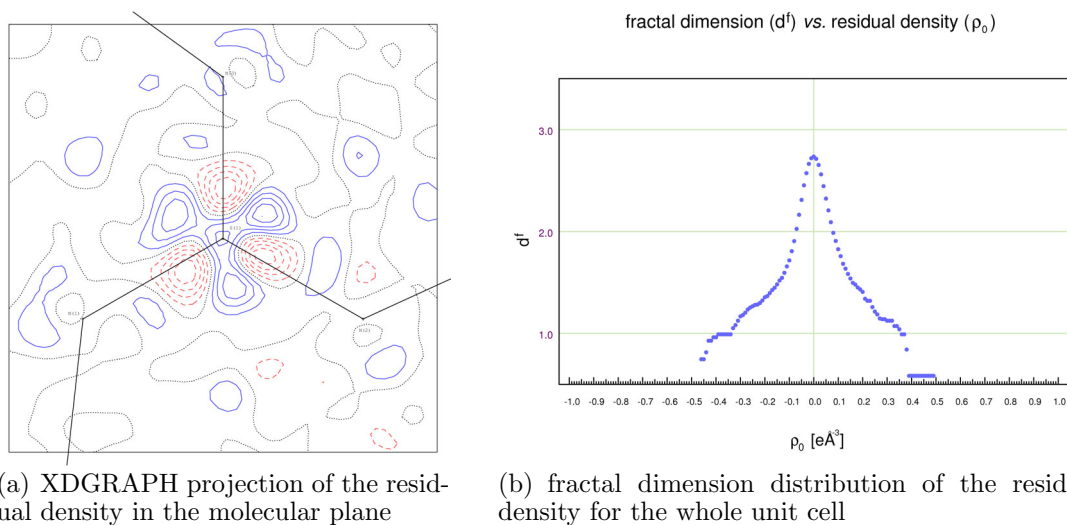


Figure 3.38: Residual density after an increase in κ' of S in $S(N^tBu)_3$ by 50%; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 e\text{\AA}^{-3}$.

In the case of a change of κ' the fractal dimension $d^f(0)$ does not give a clear hint what value of κ' is the best one. $d^f(0)$ increases slightly for increased values for κ' (at least up to 10 % increase). The highest value for $d^f(0)$ is at +10 %. The flatness

$\Delta\rho_0$ remains unchanged for an extremely wide range of κ' values. The gross residual electrons e_{gross} have their minimum at the original value.

Anomalous dispersion effects

The $\Delta f'$ and $\Delta f''$ values were changed in the SCAT table in the xd.mas file (DELF' and DELF'', respectively) and a calculation of the structure factors was performed (XDLSM with the option "cycle 0").

Both contributions to the anomalous scattering $\Delta f'$ and $\Delta f''$ were varied simultaneously and independent of each other to investigate the individual influence of the parameters on the residual density.

Simultaneous variation in $\Delta f'$ and $\Delta f''$ For a simultaneous change in $\Delta f'$ and $\Delta f''$ the respective values were changed in the SCAT section in the xd.mas file. Table 3.19 shows the results for changes from -50% to $+50\%$. Fig. 3.39 shows the residual density distribution in the plane containing the molecule and the corresponding fractal dimension distribution for an increase in $\Delta f'$ and $\Delta f''$ by 50% .

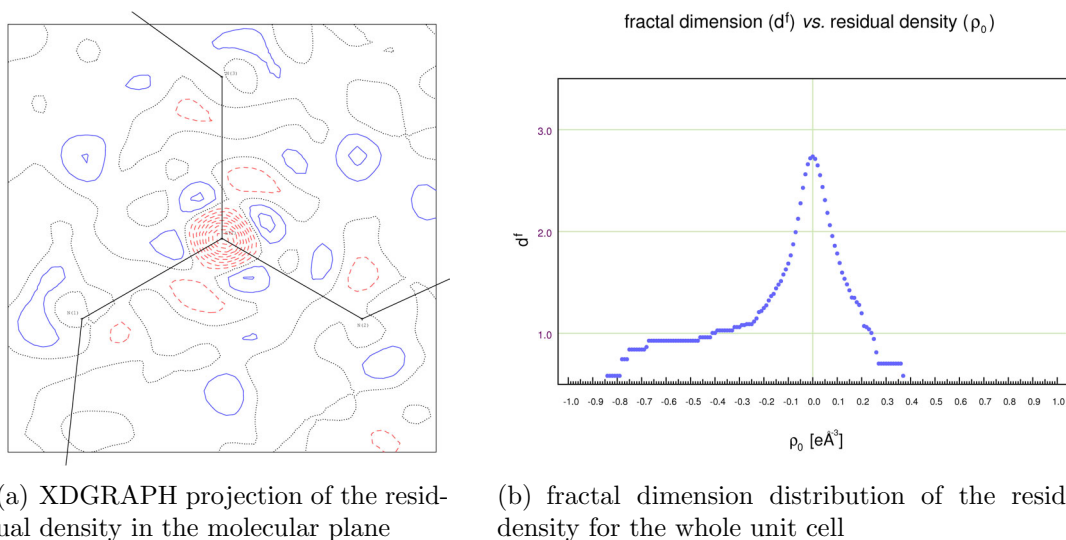


Figure 3.39: Residual density after an increase in $\Delta f'$ and $\Delta f''$ of S in $S(N^tBu)_3$ by 50% ; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 eÅ^{-3} .

	$\Delta f'/\Delta f''$	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
-50 %	0.0623/0.0617	2.7375	8.6605	-0.33	1.41	1.74
-20 %	0.0997/0.0987	2.7366	8.4933	-0.32	0.73	1.05
-10 %	0.1121/0.1111	2.7363	8.4410	-0.32	0.51	0.83
orig	0.1246/0.1234	2.7363	8.3961	-0.32	0.37	0.69
+10 %	0.1371/0.1357	2.7365	8.3595	-0.32	0.37	0.69
+20 %	0.1495/0.1481	2.7365	8.3336	-0.39	0.37	0.76
+50 %	0.1869/0.1851	2.7358	8.2911	-0.84	0.37	1.21

Table 3.19: Residual density descriptors for a change of $\Delta f'$ and $\Delta f''$ of S in experimental data of $S(N^tBu)_3$.

Change in $\Delta f'$ Table 3.20 shows the results for an individual change in $\Delta f'$ from -50 % to +50 %. Fig. 3.40 shows the residual density distribution in the plane containing the molecule and the corresponding fractal dimension distribution for an increase in $\Delta f'$ by 50 %.

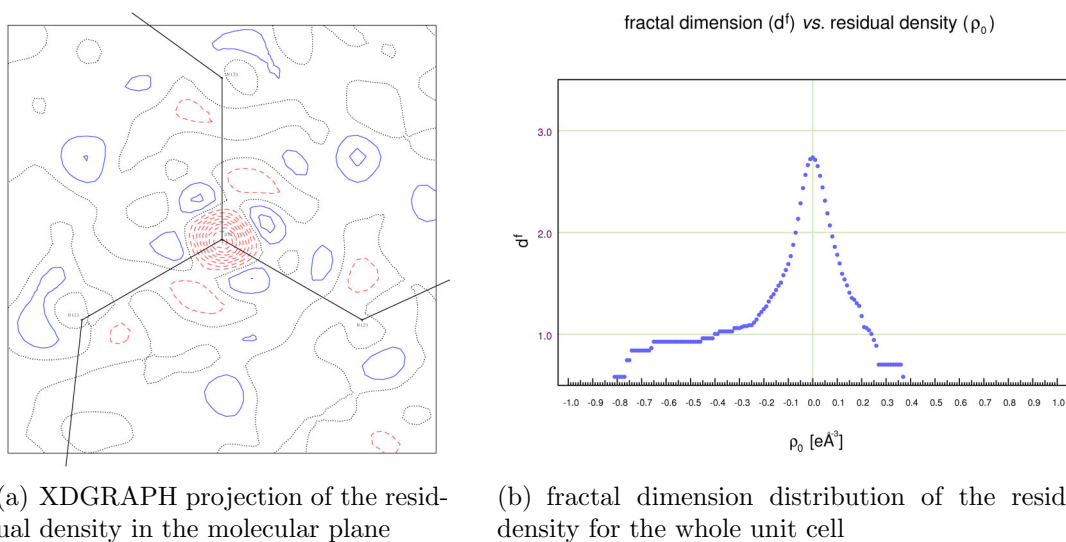


Figure 3.40: Residual density after an increase in $\Delta f'$ of S in $S(N^tBu)_3$ by 50 %; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 e\text{\AA}^{-3}$.

	$\Delta f'/\Delta f''$	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ (eÅ ⁻³)	$\rho_{0,max}$ (eÅ ⁻³)	$\Delta\rho_0$ (eÅ ⁻³)
-50 %	0.0623/0.1234	2.7374	8.5641	-0.32	1.39	1.71
-20 %	0.0997/0.1234	2.7363	8.4543	-0.32	0.72	1.04
-10 %	0.1121/0.1234	2.7364	8.4218	-0.32	0.50	0.82
orig	0.1246/0.1234	2.7363	8.3961	-0.32	0.37	0.69
+10 %	0.1371/0.1234	2.7366	8.3779	-0.32	0.37	0.69
+20 %	0.1495/0.1234	2.7365	8.3690	-0.39	0.37	0.76
+50 %	0.1869/0.1234	2.7360	8.3698	-0.81	0.37	1.18

Table 3.20: Residual density descriptors for a change of $\Delta f'$ of S in experimental data of S(N^tBu)₃.

Change in $\Delta f''$ Table 3.21 shows the results for an individual change in $\Delta f''$ from -50 % to +50 %. Fig. 3.41 shows the residual density distribution in the plane containing the molecule and the corresponding fractal dimension distribution for an increase in $\Delta f''$ by 50 %.

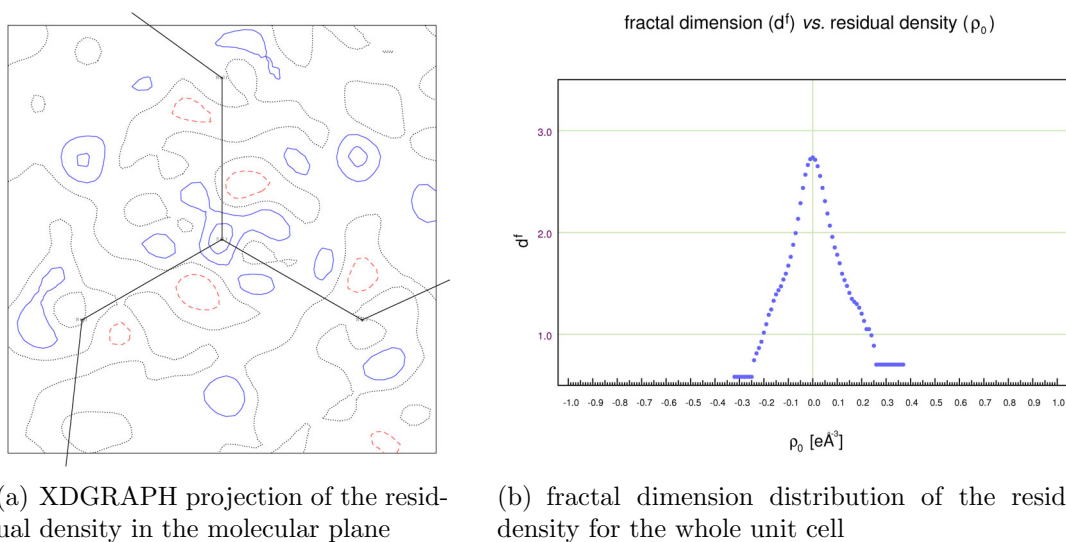


Figure 3.41: Residual density after an increase in $\Delta f''$ of S in S(N^tBu)₃ by 50 %; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 eÅ⁻³.

	$\Delta f'/\Delta f''$	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ (eÅ ⁻³)	$\rho_{0,max}$ (eÅ ⁻³)	$\Delta\rho_0$ (eÅ ⁻³)
-50 %	0.1246/0.0617	2.7364	8.4817	-0.32	0.37	0.69
-20 %	0.1246/0.0987	2.7363	8.4330	-0.32	0.37	0.69
-10 %	0.1246/0.1111	2.7364	8.4148	-0.32	0.37	0.69
orig	0.1246/0.1234	2.7363	8.3961	-0.32	0.37	0.69
+10 %	0.1246/0.1357	2.7365	8.3772	-0.32	0.37	0.69
+20 %	0.1246/0.1481	2.7364	8.3578	-0.32	0.37	0.69
+50 %	0.1246/0.1851	2.7362	8.3002	-0.32	0.37	0.69

Table 3.21: Residual density descriptors for a change of $\Delta f''$ of S in experimental data of S(N^tBu)₃.

For simultaneous as well as independent changes in $\Delta f'$ and $\Delta f''$ it was observed that all residual density descriptors, $d^f(0)$, e_{gross} and $\Delta\rho_0$, increased when $\Delta f'$ and $\Delta f''$, respectively, were decreased (see Table 3.19). For an increase of the anomalous dispersion of S up to 20 % $d^f(0)$ increases slightly, but at higher values it decreases again, whereas e_{gross} decreases steadily when $\Delta f'$ and $\Delta f''$ are increased. As expected, the flatness $\Delta\rho_0$ decreases for increasing as well as for decreasing values for $\Delta f'$ and $\Delta f''$. The increase in $d^f(0)$ for decreasing $\Delta f'$ and $\Delta f''$ as well as the decreasing value of e_{gross} for increasing $\Delta f'$ and $\Delta f''$ are quite unexpected results and are somewhat counterintuitive. The number of e_{gross} can only decrease when the average residual density decreases (see Eq. 3.8).

A closer look at the residual density file shows that this is caused by an increase in frequency of very small residual density values (± 0.02 eÅ⁻³ around 0) and a decrease of small values (± 0.3 to 0.7 eÅ⁻³ around 0). As there is no total control of the sources of error in the experiment and as different sources of error may interfere, a change in a subset of parameters might partially compensate for the non-modeled, unknown or unaccounted sources of error, thereby generating changes in the residual density, which are characteristic of the source of error instead of the changed parameters.

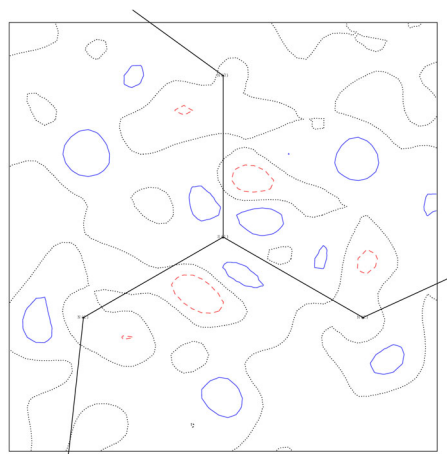
As can be seen from Table 3.9 the simulated data show the expected results: for increasing or decreasing $\Delta f'$ and $\Delta f''$ the fractal dimension $d^f(0)$ decreases which indicates a deterioration of the model, whereas the gross residual electrons and $\Delta\rho_0$ increase. One significant difference between the measured and the simulated data is a tiny, not modeled disorder.

It is remarkable that each of the parameter manipulations results in a characteristic shape in the fractal dimension distribution. This is a big advantage compared to $\Delta\rho_0$ which cannot give any information about the source of error but only the

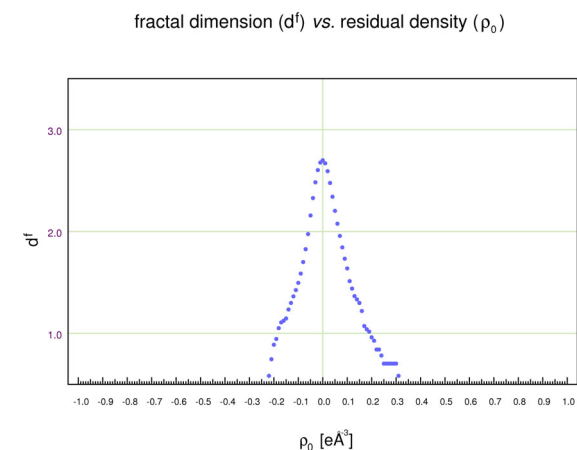
width of the residual density distribution. Thus, $d^f(0)$ is much more appropriate to describe model inadequacies and features than $\Delta\rho_0$.

3.5.2.3 Influence of the truncation of the experimental resolution

To investigate the electronic situation in a molecule an X-ray experiment is performed at high resolution to analyze fine details in the atomic properties. Unfortunately, the signal to noise ratio decreases with increasing resolution. Furthermore, the number of reflections increases rapidly with increasing resolution (as a rule of thumb: $80/d^3$ unique reflections per non-hydrogen atom for organic centrosymmetric structures and $40/d^3$ reflections for non-centrosymmetric structures, respectively). This leads to residual densities being more and more affected by noise with increasing resolution. Thus, it has become common habit to employ the whole resolution for the refinement, but to truncate the data for creating the Fourier maps. As a consequence, not only the high-frequency noise is filtered out but presumably also the discrepancy between the model and the high-order data. To determine the influence of the truncation in resolution (on the residual density) several residual density grids of the fully refined model for $S(N^tBu)_3$ with resolution cutoffs at 1.14 \AA^{-1} , 1.00 \AA^{-1} and 0.80 \AA^{-1} were generated and analyzed with jnk2RDA. The grids were calculated with XDFOUR for the Fourier backtransformation in the XDFOUR section in the xd.mas file. The original resolution of the data set is 1.14 \AA^{-1} . Figs. 3.42 and 3.43 show residual density plots for the truncation at 1.00 \AA^{-1} and 0.80 \AA^{-1} , respectively. Table 3.22 shows the corresponding values for the residual density descriptors.

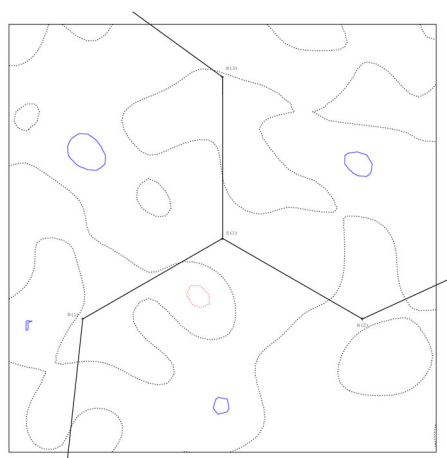


(a) XDGRAPH projection of the residual density in the molecular plane

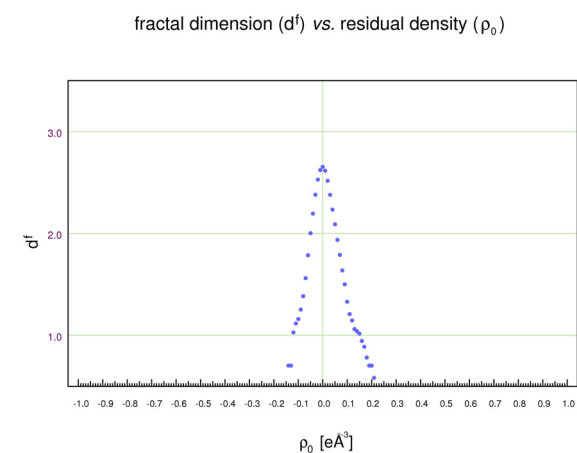


(b) fractal dimension distribution of the residual density for the whole unit cell

Figure 3.42: Residual density after application of a resolution cutoff for the Fourier synthesis at $(\sin \theta/\lambda)_{max} = 1.00 \text{ \AA}^{-1}$ in $S(N^tBu)_3$; for a comparison with the untruncated data see Fig. 3.34; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 e\AA^{-3} .



(a) XDGRAPH projection of the residual density in the molecular plane



(b) fractal dimension distribution of the residual density for the whole unit cell

Figure 3.43: Residual density after application of a cutoff for the Fourier synthesis at $(\sin \theta/\lambda)_{max} = 0.80 \text{ \AA}^{-1}$ in $S(N^tBu)_3$; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: 0.1 e\AA^{-3} .

cutoff at resolution $(\sin \theta/\lambda)_{max}$ (\AA^{-1})	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
no cutoff (1.14)	2.7366	8.3851	-0.32	0.37	0.69
1.00	2.6989	7.2521	-0.23	0.32	0.55
0.80	2.6543	6.2673	-0.15	0.21	0.36

Table 3.22: Applied cutoff in the resolution for the Fourier grid in $S(N^tBu)_3$.

The residual density distribution calculated from a truncated resolution at 1.00\AA^{-1} (Fig. 3.42) shows significantly decreased features (compared with Fig. 3.34). e_{gross} decreases from $8.3851 e$ to $7.2521 e$ and $\Delta\rho_0$ decreases from $0.69 e\text{\AA}^{-3}$ to $0.55 e\text{\AA}^{-3}$. The features in the main plane of the molecule (Fig. 3.42) get much smoother and the width of the fractal dimension distribution gets smaller. But on the other hand the fractal dimension $d^f(0)$ decreases from 2.7366 to 2.6989. As the model parameters were not adjusted this exclusively reflects the truncation of the high-order components. Compared to an increase in $d^f(0)$ of 0.0241 from the IAM to the fully refined multipole model (see Table 3.12) the loss in $d^f(0)$ due to Fourier truncation is substantial and even overcompensates the gain from the refinement. Also, that the gross residual electrons decrease without a change in the model shows that this is only a phony improvement.

The differences in the absolute values for $d^f(0)$ in Tables 3.12 and 3.22 result from the fact that no $I/\sigma(I)$ cutoff was applied here. A refinement against data with $I > 3\sigma(I)$ is comparable to a refinement against data truncated in the resolution, as the $I/\sigma(I)$ ratio decreases for increasing resolution. As was shown in section 3.5.1.1 this leads to lower values for $d^f(0)$.

For a truncation at even lower resolution (0.80\AA^{-1}) the features disappear almost completely (compare Figs. 3.34 and 3.43). e_{gross} decreases to $6.2643 e$ and $\Delta\rho_0$ to $0.36 e\text{\AA}^{-3}$. But again the fractal dimension $d^f(0)$ decreases to 2.6543 which shows that no improvement in the model was achieved.

For consistency reasons the residual density should always be described for the same resolution as the model and its corresponding parameters. It is inconsistent to refine a density model against the whole data range and to show residual density plots based on a truncated data set.

3.5.2.4 Extinction effects

Extinction occurs for crystals of high quality and when the mosaicity of the crystal is relatively low. Then the scattered beam is less divergent than it would be with high

mosaicity and can serve itself as a primary beam which can cause further diffraction effects. Extinction is observed especially at low angles for strong reflections. There are two different extinction effects, primary and secondary extinction. Primary extinction means that the diffracted beam itself can serve as a primary beam for further diffraction (Fig. 3.44(a)). Secondary extinction means that the incident beam is weakened by strong diffraction at the first lattice planes such that only a weakened beam reaches deeper planes. This reflection thus loses intensity already before being diffracted by all planes (Fig. 3.44(b)). Secondary extinction is assumed to play the major role but as the two forms of extinction cannot be distinguished easily both are corrected simultaneously with an empirical extinction correction factor χ . In SHELXL this is done by

$$F_{calc}(corr) = k \frac{F_{calc}}{(1 + 0.001 \cdot \chi \cdot F_{calc}^2 \cdot \lambda^3 / \sin 2\theta)^{\frac{1}{4}}} \quad (3.23)$$

with k = scale factor, χ = extinction coefficient, λ = wavelength, θ = diffraction angle.

With Cu radiation extinction occurs more often because of the higher intensity yield.

Shock-cooling of the crystal (*e.g.* to 100 K) reduces the probability of extinction as the mosaicity is increased.

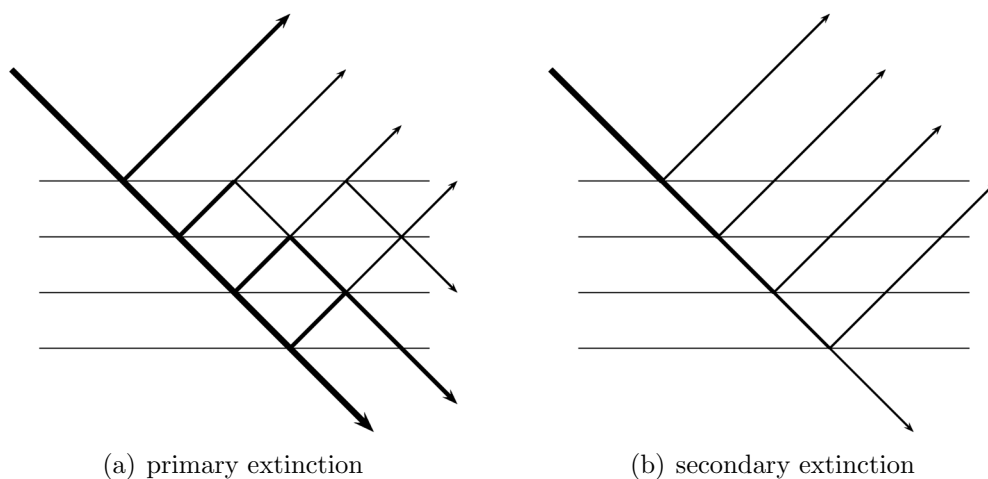


Figure 3.44: Illustration of primary and secondary extinction.

To investigate the effects of extinction on the residual density two “*.hkl” files from the IAM refinement of bullvalene triepoxide ($C_{10}H_{10}O_3$)^[54] were generated with SHELXL: one which was corrected for extinction and another one which was not. The option “list 6” together with “exti” in SHELXL generates a “*.fcf” file with corrected intensities. Without giving the option “exti”, “list 6” produces a “*.fcf”

file with no correction for extinction. The two reflection files were converted into a SHELX format “*.hkl” file and imported into XD with XDINI. A least-squares refinement against F^2 with ten cycles of XDLSM (MODEL -2 2 1 0) was performed for each of the two test sets. XDFOUR was used to generate residual density grid files for the whole unit cell for subsequent analysis with jnk2RDA.

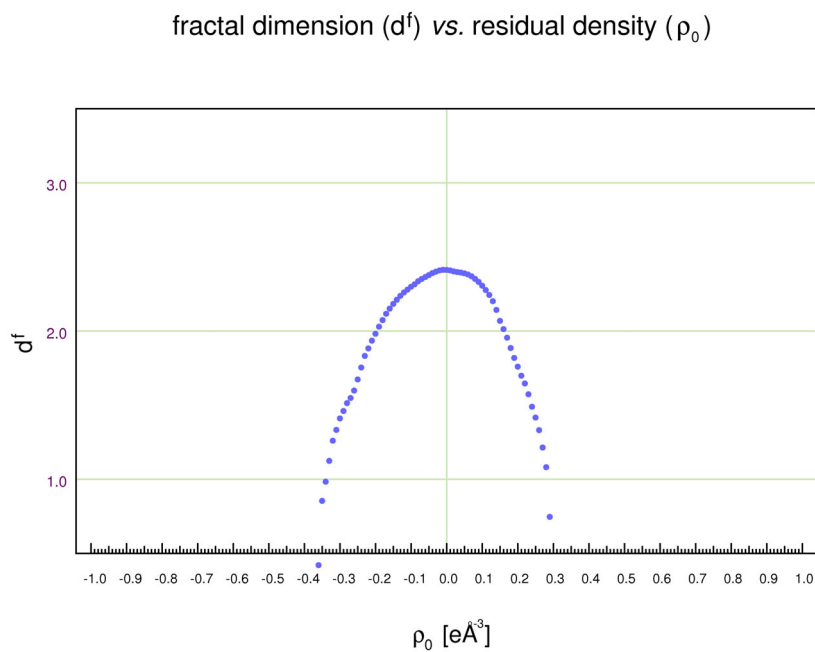


Figure 3.45: Fractal dimension distribution of the residual density of bullvalene trisepoxide without extinction correction.

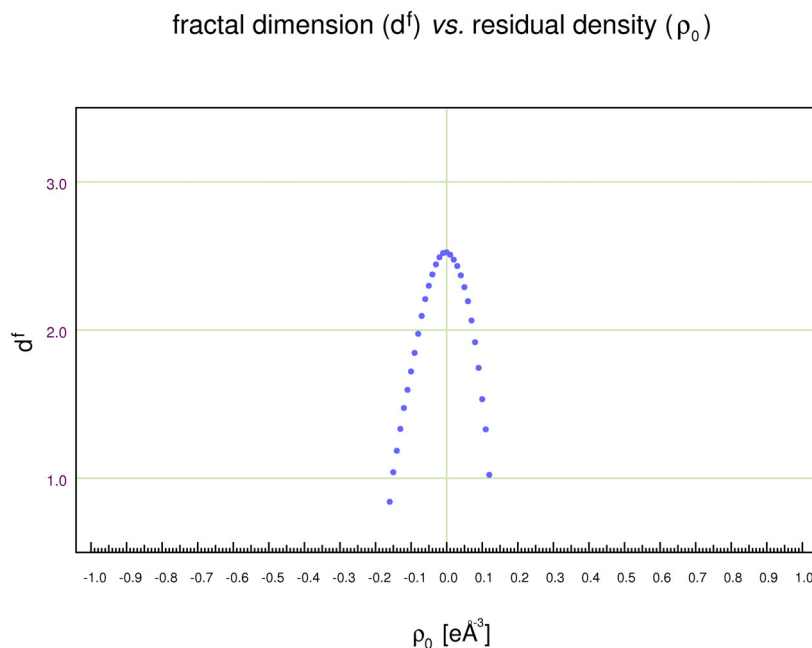


Figure 3.46: Fractal dimension distribution of the residual density of bullvalene trisepoxide with extinction correction.

correction for extinction	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
off	2.4129	30.2552	-0.36	0.29	0.65
on	2.5228	11.0321	-0.16	0.12	0.28

Table 3.23: Residual-density descriptors applied to experimental data for bullvalene trisepoxide to study the effects of the extinction correction as implemented in SHELXL.

As can be seen from Figs. 3.45 and 3.46 and Table 3.23 the extinction correction has a pronounced effect on the residual density although the extinction coefficient χ for bullvalene trisepoxide was refined by SHELXL to be only 0.04935(240). The gross residual electrons decrease from 30.2552 e to 11.0321 e, the fractal dimension $d^f(0)$ increases from 2.4129 to 2.5228 and the residual density gets much flatter (decrease in $\Delta\rho_0$ from $0.65 e\text{\AA}^{-3}$ to $0.28 e\text{\AA}^{-3}$). No such distinct change in the descriptors was observed for the adjustment of any other single parameter nor for the total difference between IAM and MM.

3.5.2.5 Grid size effects

According to Eq. 3.5 the fractal dimension is only defined in the limit $\varepsilon \rightarrow 0$. This definition, however, refers to mathematical fractals, only, which exist in unlimited

detail. In contrast, natural fractals like coastlines or the residual density distribution exist only on limited scales. For example, the shortest possible ruler for measuring the length of the coastline is an atom, which has a finite size. In analogy, the experimentally accessible details of the residual density distribution are limited by the experimental resolution. In consequence, it is not necessary to calculate $d^f(0)$ for unlimited spatial resolution but it is not *a priori* clear which spatial resolution is sufficient. Therefore, the influence of the size of the residual density grid on the values for d^f was examined. However, there are limits as the grid files which are written by XDFOUR become quite large for increasing grid size. A $50 \times 50 \times 50$ grid has about 2MB, a $100 \times 100 \times 100$ grid has a size of about 17MB and a $1000 \times 1000 \times 1000$ grid has already a size of 17 GB. For a typical unit cell of a small molecule structure of size $a = 10 \text{ \AA}$, $b = 10 \text{ \AA}$, $c = 10 \text{ \AA}$ the $1000 \times 1000 \times 1000$ grid corresponds to a distance of 0.01 \AA between the grid points, which is far from being an infinitesimal small grid spacing but on the other hand it is only a small fraction of the resolution d .

Again, the fully refined model (refined against experimental data) of $\text{S}(\text{N}^t\text{Bu})_3$ was used for the test of the grid size effect. Six grids were generated with XDFOUR with 50^3 , 100^3 , 200^3 , 500^3 , 750^3 and 1000^3 grid points, respectively. As can be seen from Fig. 3.47 and Table 3.24 with increasing grid size $d^f(0)$ decreases steadily while $\Delta\rho_0$ increases. This is in analogy to the decreasing fractal dimension of the coastline of Britain with decreasing edge length in the box-counting algorithm (compare with Fig. 3.15 and the related text): more features in the coastline were observed with decreasing size of the ruler. The same is valid for the residual density: with increasing resolution of the grid the calculated values describe the features in greater detail than a coarse grid with averaged values of the residual density. As more features appear with a finer grid the minimum and maximum values in the residual density ($\rho_{0,min}$, $\rho_{0,max}$) can be determined more reliably. This is the reason why $\Delta\rho_0$ increases from 0.69 e\AA^{-3} for the $50 \times 50 \times 50$ grid to 0.78 e\AA^{-3} for the $1000 \times 1000 \times 1000$ grid. $d^f(0)$ decreases from 2.7367 for the $50 \times 50 \times 50$ grid to 2.4445 for the $1000 \times 1000 \times 1000$ grid. The gross residual electrons are virtually independent on the grid size. They remain stable for different grid sizes with a value of 8.38 e. As the value for $\Delta\rho_0$ does not change anymore for an increase in grid size from the $500 \times 500 \times 500$ grid to the $1000 \times 1000 \times 1000$ grid ($\Delta\rho_0 = 0.78 \text{ e\AA}^{-3}$) the residual density values seem to be converged up to the second decimal place. But the fractal dimension $d^f(0)$ is still not converged even for the $1000 \times 1000 \times 1000$ grid. As the file with 1000^3 grid points has already a size of 17 GB it is not very efficient to calculate even larger grids. In view of the definition of the fractal dimension (Eq. 3.5) the slow convergence is expected

due to the logarithm involved. Moreover, the residual density values themselves still change with the grid size which prevents a faster convergence. However, for a comparison of different models only the relative changes are of importance. Finally, for a limited resolution in reciprocal space it seems inappropriate to enforce the limit $\varepsilon \rightarrow 0$ in real space.

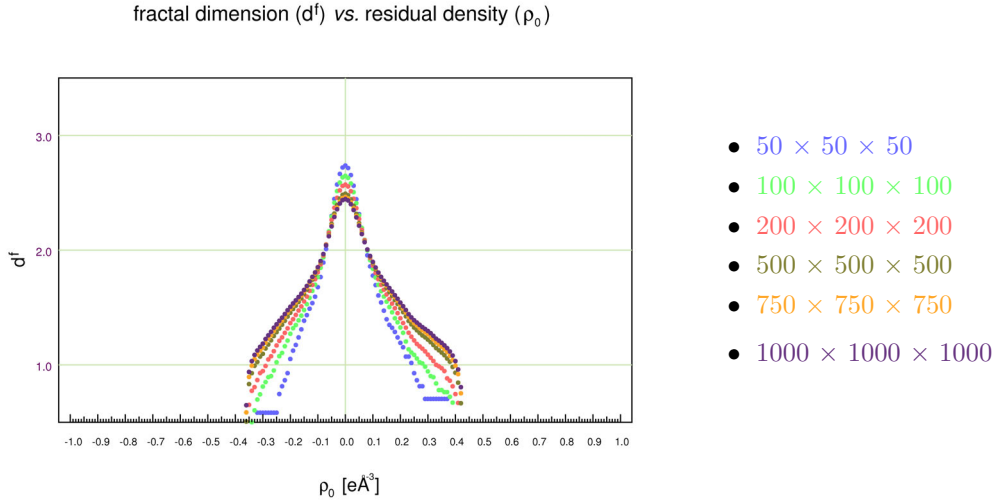


Figure 3.47: Fractal dimension for different grid sizes; the first three grids and the last three grids each form a block; within the first block ρ_0 converges, whereas within the second block the convergence of $d^f(0)$ becomes visible.

$n_x = n_y = n_z$	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}(d^f(\rho_0) = 2)$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}(d^f(\rho_0) = 2)$ ($e\text{\AA}^{-3}$)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
50	2.7367	8.3849	-0.0708	0.0761	-0.32	0.37	0.69
100	2.6472	8.3794	-0.0741	0.0796	-0.34	0.39	0.73
200	2.5703	8.3794	-0.0751	0.0807	-0.35	0.41	0.76
500	2.4913	8.3794	-0.0754	0.0809	-0.36	0.42	0.78
750	2.4628	8.3794	-0.0754	0.0810	-0.36	0.42	0.78
1000	2.4445	8.3794	-0.0754	0.0810	-0.36	0.42	0.78

Table 3.24: Residual density descriptors for a change of the grid size in $S(N^t\text{Bu})_3$.

An interesting detail of Fig. 3.47 is that the values of $\rho_{0,min}$ and $\rho_{0,max}$ for $d^f(\rho_0) = 2$ remain surprisingly stable for the different grid sizes. They deviate by less than $0.005 e\text{\AA}^{-3}$ (Table 3.24) although $d^f(0)$ and $\Delta\rho_0$ change considerably. By the averaging process taking place in the calculation of the residual density for spatially limited subvolumes of the unit cell, the frequency of small residual density values is overestimated whereas the frequency of large residual density values is underestimated. This leads to the characteristic behavior in $\Delta\rho_0$ and in $d^f(0)$. The

residual density values for which $d^f(\rho_0) = 2$ is valid mark the border between under- and overestimation.

3.5.2.6 Grid resolution effects

For the calculation of the residual density grid with XDFOUR the user is free in the choice of the size of the grid. The number of grid points along the axes of the desired plane or cuboid can be chosen independent of each other. As was seen in the previous section, for a given set of structure factors \mathbf{F}_{obs} and \mathbf{F}_{calc} the value of $d^f(0)$ decreases for increasing grid size. In other words: $d^f(0)$ is always overestimated for finite ε . For a given fixed number of the grid size $N(\varepsilon) = n_x \times n_y \times n_z$, different choices of n_x , n_y and n_z can be made leading to the same grid size $N(\varepsilon)$ but representing different spatial resolutions. Therefore, the combination of n_x , n_y and n_z leading to the lowest value for $d^f(0)$ is the most efficient choice. Thus, several grids were calculated with XDFOUR with values for n_x , n_y and n_z being in proportion to the cell axes a , b and c . For this study a different structure than $S(N^tBu)_3$ was chosen because the cell parameters of $S(N^tBu)_3$ were not very different from each other ($a = 9.3228 \text{ \AA}$, $b = 9.3455 \text{ \AA}$, $c = 10.6675 \text{ \AA}$) which makes it difficult to see an effect of the permutation of n_x , n_y and n_z . Octamethylcyclotetrasilazane [$c\text{-}\{Me_2Si-N(H)\}_4$] ^[55] (OMCTS) crystallizes in the monoclinic space group $P2/c$ with the cell constants $a = 11.5216 \text{ \AA}$, $b = 6.3499 \text{ \AA}$, $c = 22.6282 \text{ \AA}$, $\beta = 90.474^\circ$. Residual density grids with varying values for n_x , n_y and n_z were calculated with XDFOUR for the fully refined multipole model of OMCTS. In proportion to the cell constants the chosen values were $x = 115$, $y = 63$ and $z = 226$. Besides the grids with different sizes in n_x , n_y and n_z also the grids with equal sizes were calculated.

n_x, n_y, n_z	$N(\varepsilon)$	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($e\text{\AA}^{-3}$)	$\rho_{0,max}$ ($e\text{\AA}^{-3}$)	$\Delta\rho_0$ ($e\text{\AA}^{-3}$)
zzz	11 543 176	2.5368	29.1488	-0.25	0.26	0.51
xyz	1 637 370	2.5722	29.1485	-0.25	0.25	0.50
yxz	1 637 370	2.5902	29.1488	-0.24	0.26	0.50
zyx	1 637 370	2.5983	29.1522	-0.25	0.25	0.50
xxx	1 520 875	2.5989	29.1511	-0.25	0.25	0.50
zxy	1 637 370	2.6230	29.1379	-0.24	0.24	0.48
yzx	1 637 370	2.6249	29.1489	-0.24	0.25	0.49
xzy	1 637 370	2.6330	29.1385	-0.24	0.24	0.48
yyy	250 047	2.6457	29.1411	-0.24	0.24	0.48

Table 3.25: Residual density descriptors for a varying grid resolution for OMCTS.

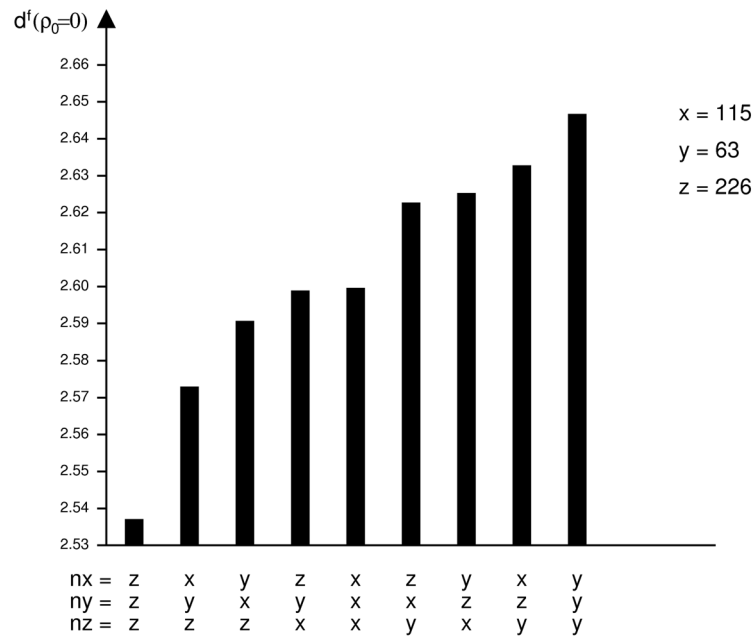


Figure 3.48: Fractal dimension $d^f(0)$ for a variation of the grid resolution; the grid sizes n_x, n_y, n_z are varied proportional to the cell axes a, b, c .

A comparison of the three different grids with identical numbers for n_x, n_y and n_z ($n_x = n_y = n_z = 63, 115$ and 226 , respectively) shows the same result as was seen in the previous section, namely the decrease of $d^f(0)$ for increasing $N(\varepsilon)$ (2.6457 for $n_x = n_y = n_z = 63$ and 2.5368 for $n_x = n_y = n_z = 226$) accompanied by an increase in $\Delta\rho_0$ ($0.48 \text{ e}\mathring{\text{A}}^{-3}$ for $n_x = n_y = n_z = 63$ and $0.51 \text{ e}\mathring{\text{A}}^{-3}$ for $n_x = n_y = n_z = 226$). But the difference in overall grid points and thus the difference in computing time and memory space is quite large ($11\,543\,176$ grid points and a file size of 201 MB for the 226^3 grid and $250\,047$ grid points and a file size of 4.37 MB for the 63^3 grid). The calculation and analysis of a grid with a spatial resolution in proportion to the unit cell axes turned out to be much more efficient. The grid with $n_x = 115, n_y = 63$ and $n_z = 226$ is a good compromise between computing efforts and convergence requirement: it gives the second smallest value for the fractal dimension in total and the smallest value for the grids with constant $N(\varepsilon)$ ($d^f(0) = 2.5722$) and has only a file size of 28.6 MB with a moderate number of grid points of $N(\varepsilon) = 1\,637\,370$.

3.5.3 Residual density descriptors at the density model and parameter optimum

A topic that has to be dealt with is the expected values of the residual density descriptors at the optimum. From a phenomenological point of view, the descriptors in most cases assumed extremal values for the optimum parameters in the application

to theoretical and experimental data. This tendency is particularly distinct in the application to theoretical data, whereas in the application to experimental data it is less obvious. As mentioned earlier, the experimental data are influenced by errors, not all of which might be known or even controlled. One unconsidered error in the experimental data is *e.g.* the small disorder. The interference of different sources of error can lead to unexpected values for the residual density descriptors.

It is expected that $d^f(0)$ assumes its maximum or minimum for the parameter optimum, depending on whether the source of error acting effectively leads to the artificial construction or destruction of zero residual density values. Most of the parameters from the multipole model belong to the first group. Therefore, as a rule of thumb, it can be stated that a maximum value in $d^f(0)$ is expected to characterize the optimum multipole model.

e_{gross} is expected to be at its minimum value for the model and parameter optimum in exactly the same sense in which the usual agreement factors are expected to be at a minimum. The comparison with the agreement factors is vital, because it is well known that the R -value can be lowered artificially at the expense of the physical or chemical meaning of the model or by *e.g.* data truncation. The same holds for e_{gross} .

The only exception for e_{gross} being at its minimum is in the simultaneous and separate variation of $\Delta f'$ and $\Delta f''$ (for the sulfur atom) for experimental data, exclusively. It is assumed that this exceptional behavior possibly can be traced back to the non-modeled disorder or to further sources of error. This assumption is suggested by the expected behavior for theoretical data, in which full control of all sources of error can be anticipated.

3.5.4 Analysis of the refinement of a disordered and non-centrosymmetric structure

Usually, it is very difficult to refine a disordered structure with multipole methods because of the high correlation between parameters such as positional and thermal motion parameters with the monopole- and multipole population parameters. These problems occur especially when the disorder is such that the two (or more) disordered entities share the same positional parameters, as is the case in *N*-phenylpyrrole. *N*-phenylpyrrole $\text{H}_4\text{C}_4\text{N}(\text{C}_6\text{H}_5)$ (PP) (Fig. 3.49) crystallizes in the non-centrosymmetric orthorhombic space group $P2_12_12_1$ with the cell axes $a = 5.6405 \text{ \AA}$, $b = 7.5996 \text{ \AA}$, $c = 17.6532 \text{ \AA}$. There is one molecule in the asymmetric unit which is disordered by a 2-fold rotation axis. The occupation factors for the two disordered positions were 90% and 10%, respectively.

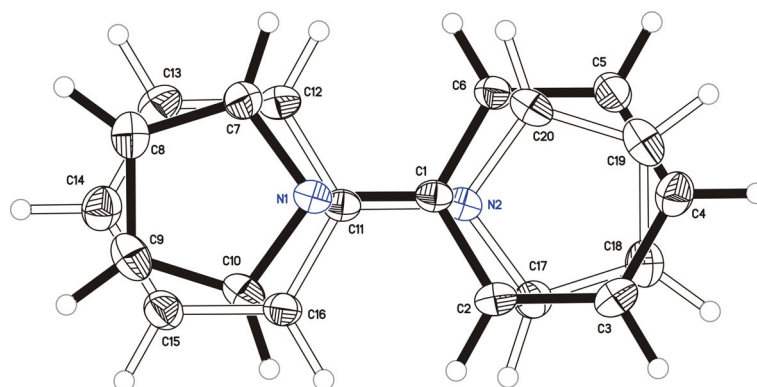


Figure 3.49: *N*-phenylpyrrole (PP) with the two disordered positions of the molecule. The main domain (occupation factor 90 %) is plotted in black, the minor domain (10 %) in white. The anisotropic displacement parameters are shown on the 50 % probability level.

The static disorder in PP was refined with multipole methods [56]. To prove the reliability of the disorder modeling the residual densities for both models (the one which does not take the disorder into account and the one which does) were analyzed with jnk2RDA. For this purpose two residual density grids were calculated with XDFOUR. Fig. 3.50 shows the results for the model where only the main domain was refined.

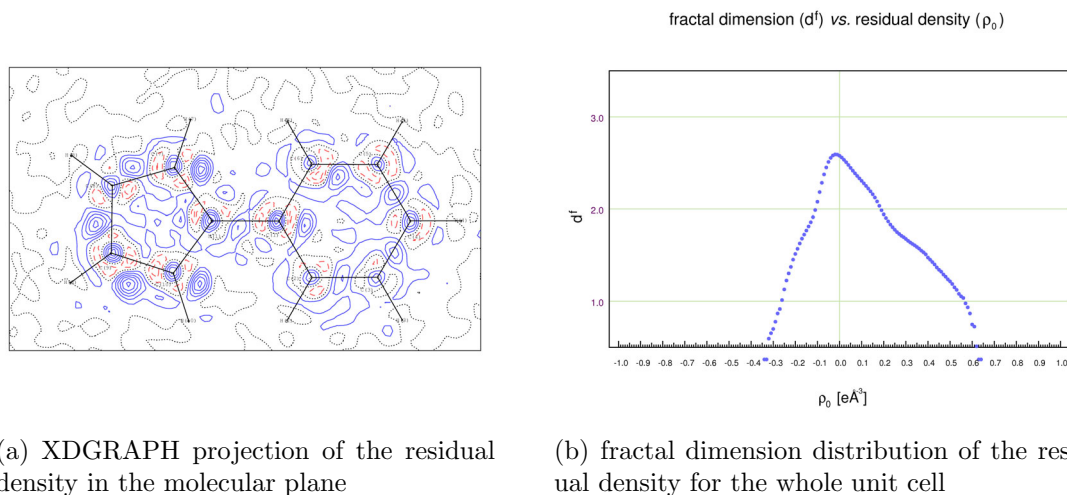


Figure 3.50: Residual density of PP without refinement of the disorder; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

In Fig. 3.50(a) the residual density in the plane of the molecule is shown. The grid was calculated in a plane of $10 \text{ \AA} \cdot 6 \text{ \AA}$ with a resolution of 100×60 grid points. In Fig. 3.50(b) the corresponding fractal dimension distribution of the residual density for the whole unit cell (with a resolution of $56 \times 76 \times 176$ grid points) is plotted. It can be clearly seen from both figures that the residual density is far from being

flat and featureless. The unmodeled positions for the second domain are clearly visible as positive residual density (blue solid lines) between the atomic positions of the modeled domain and as negative residual density (red dashed lines) around the atomic positions. The positive residual density at the atomic positions of the first domain come from the anisotropic displacement parameters that were determined as too high due to the second position of the molecule that was not refined. The positive residual density of the second domain is compensated for by increased values for U_{ij} . This behavior was already described in section 3.5.1.3 and can also be seen from the fractal dimension distribution. The unmodeled second domain of PP shows up as a shoulder in the positive residual density region, whereas the main domain, which is refined with a too high occupation factor, contributes to the shoulder in the negative residual density region. The fractal dimension $d^f(0)$ is 2.5770, e_{gross} is 16.0809 e and $\Delta\rho_0$ is $0.98 \text{ e}\text{\AA}^{-3}$.

refinement of disorder	$d^f(0)$	e_{gross} (e)	$\rho_{0,min}$ ($\text{e}\text{\AA}^{-3}$)	$\rho_{0,max}$ ($\text{e}\text{\AA}^{-3}$)	$\Delta\rho_0$ ($\text{e}\text{\AA}^{-3}$)
no	2.5770	16.0809	-0.34	0.64	0.98
yes	2.6286	6.7356	-0.12	0.15	0.27

Table 3.26: Residual density descriptors for *N*-phenylpyrrole with and without the refinement of the disorder.

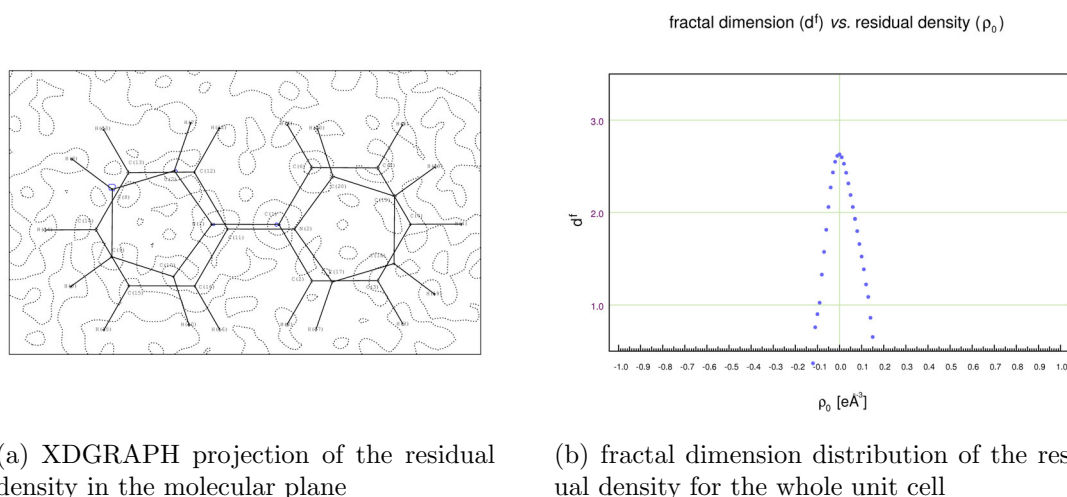


Figure 3.51: Residual density of PP with refinement of the disorder; (a) blue solid lines: positive residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

In Fig. 3.51 the residual density in the plane containing the two domains of PP (Fig. 3.51(a)) and the corresponding residual density distribution for the whole unit

cell (Fig. 3.51(b)) are shown. The fractal dimension $d^f(0)$ increased to 2.6286, e_{gross} decreased to 6.7356 e and $\Delta\rho_0$ decreased to $0.27 \text{ e}\text{\AA}^{-3}$. The residual density descriptors show that the features in the residual density decreased not only in the region around the molecule but in the whole unit cell (see also Table 3.26). The refinement of the second disordered position perfectly describes the electron density in the unit cell. No distinct features are present in the residual density anymore. Although the refinement of disordered molecules with multipole methods is usually avoided due to possible correlations in the parameters sometimes no other crystal is available and the crystallographer has no other choice. In this case, however, the residual density analysis showed that it is possible to refine a disordered structure successfully with the multipole method.

3.5.5 Correction for negative intensity observations

One of the several problems that can occur with an X-ray measurement is the presence of negative intensity observations. This happens when the background is measured to be larger than the signal, *e.g.* due to counting statistics. This is of particular importance for weak reflections. For the refinement these reflections are normally omitted or set to zero which introduces an error. Usually, the number of negative intensity observations becomes larger for higher resolution.

French and Wilson ^[23] have introduced a method to overcome this problem for protein structures. Due to the large unit cells, there is usually an abundance of reflections even at low resolution for protein structures. This renders them ideal for statistical approaches. French and Wilson compare the actual intensity distribution from the measurement including negative intensity observations to an idealized non-negative intensity distribution, which was determined once for the centrosymmetric and for the non-centrosymmetric case ^[57]. Several assumptions, such as homoatomicity and independence as well as uniformity of atomic density distributions, underly the derivation of the idealized distributions. The space group is not taken into account. By employing Bayesian statistics the measured intensities are corrected for the prior knowledge that the intensity must be non-negative and that its probability density distribution should be reminiscent of the idealized distribution. In this correction procedure, the negative intensity observations are shifted to positive values, whereas observations of strong positive reflections remain almost unchanged. As a result a reflection file with non-negative intensity observations only is obtained.

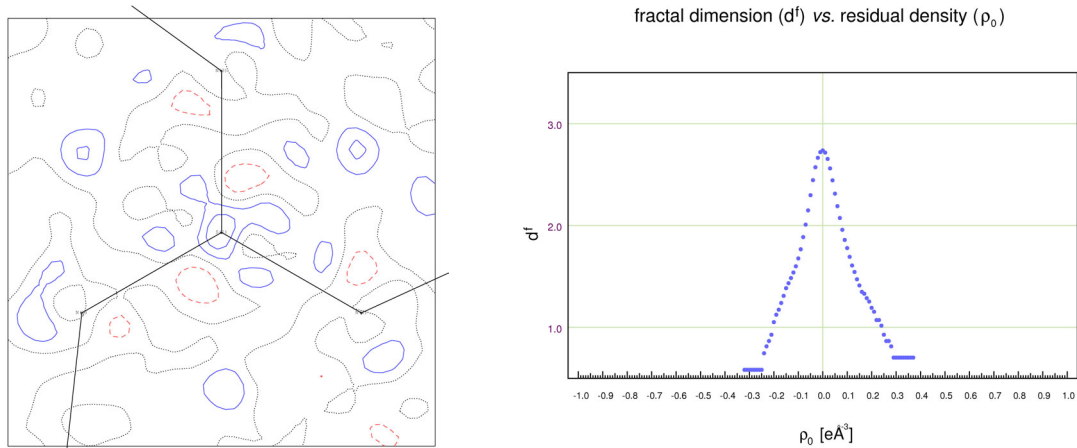
For high resolution structures where data are available up to a resolution exceeding the atomic resolution by far (*e.g.* $d = 0.5 \text{ \AA}$) an algorithm was developed and implemented into a program called `histomatch_fco` (see section 5) which uses the his-

togrammatic distribution of the observed and the calculated structure factors F_{obs} and F_{calc} for the adjustment of negative intensity observations.

Histogram matching is a method widely used in protein crystallography for density modification and phase improvement [58]. In histogram matching the frequency distribution of the electron density *versus* the electron density values is plotted in a histogram. Surprisingly, the shape of this distribution depends only on the resolution and the temperature factor but not on the content of the unit cell. Thus, the distribution of the electron density resulting from phases calculated from *e.g.* a Multiple Isomorphous Replacement (MIR) experiment can be fitted with an electron density distribution calculated *e.g.* from atomic positions from any other protein molecule if only the molecular region is taken [59] or from a protein structure with similar size of the unit cell and a similar number of atoms [60] provided the resolution is the same. The fitted electron density is more exact than the original one and so are the resulting phases.

In the program `histomatch_fco` the observed intensities F_{obs}^2 are sorted in descending order and are stored together with their corresponding hkl indices. The calculated intensities F_{calc}^2 are sorted accordingly. Then, the F_{obs}^2 , which comprise negative intensity observations, are substituted by the F_{calc}^2 , which are only positive, whereas the set of hkl is kept fixed. The distribution of the intensities remains the same, *i.e.* the largest F_{obs}^2 is replaced by the largest F_{calc}^2 , the second largest F_{obs}^2 with the second largest F_{calc}^2 , and so on. It might be expected that this results just in a replacement of the F_{obs}^2 by their corresponding F_{calc}^2 and thus simply equates to an exchange of the measured parameters by the calculated ones. But it turned out that only about 1% of the F_{obs}^2 are replaced by their corresponding F_{calc}^2 .

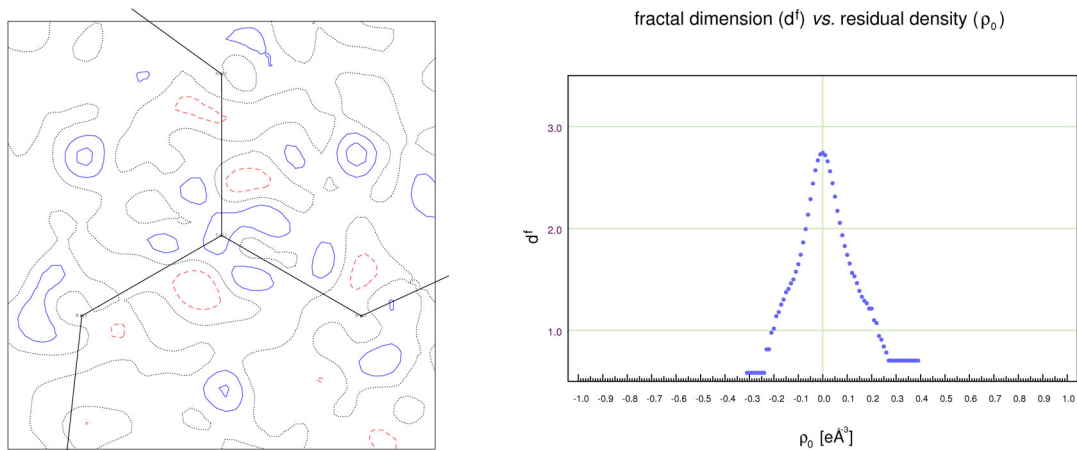
The intensity file of $S(N^tBu)_3$ has been corrected for negative intensities with the program `histomatch_fco`. The number of F_{obs}^2 that have been replaced by their F_{calc}^2 (the ones with the same hkl indices) was 108, which is 0.59%, only. A subsequent refinement against the new data was performed with XDLSM (with the option “sigobs 0”).



(a) XDGRAPH projection of the residual density in the molecular plane

(b) fractal dimension distribution of the residual density for the whole unit cell

Figure 3.52: Residual density before application of the intensity correction; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.



(a) XDGRAPH projection of the residual density in the molecular plane

(b) fractal dimension distribution of the residual density for the whole unit cell

Figure 3.53: Residual density after application of the intensity correction; (a) blue solid lines: positive residual density, red dashed lines: negative residual density, gray dotted lines: zero residual density, contour spacing: $0.1 \text{ e}\text{\AA}^{-3}$.

The R_2 -value decreased from 2.32% to 2.10%, the number of data used in the refinement increased from 17520 to 18250 as there were no negative intensities anymore. $d^f(0)$ increased from 2.7366 to 2.7423, e_{gross} decreased from 8.3851e to 8.2609e and $\Delta\rho_0$ remained the same ($0.71 \text{ e}\text{\AA}^{-3}$). Interestingly, the features in the residual density did not disappear. This can be seen mainly from Fig. 3.53 as the positive residual density that stems from the unrefined disorder (the blue lines

between each pair of nitrogen atoms, generated by rotation of 60° about the sulfur atom). The comparison with the residual density before the histogram matching (Fig. 3.52) shows that the presence and even the intensity of the features have almost not changed.

4 Labyrinthopeptin A2

4.1 Introduction

Since the discovery of penicillin in 1928 by Alexander Fleming antibiotics have gained much importance in the medical area. With their help many infectious illnesses could be defeated and without antibiotics many people would have died and still would die from bacterial diseases. Antibiotics are metabolites produced primarily by bacteria or fungi to inhibit the growth of other microorganisms or even to kill them. This is a natural defense against competitors for food and nutrients. Antibiotics can act in three different modes of action: bacteriostatic, *i.e.* they prevent the growth of the bacteria, bactericidal, *i.e.* the bacteria are killed, or bacteriolytic, *i.e.* the bacteria are killed and their cell wall is destroyed. Antibiotics are among the most widely prescribed drugs. From about 8000 up to date known antibiotics only 1% are used in medical treatment. Antibiotics are very diverse in their structures and thus also in their mechanisms to defeat bacteria. Some of the most important antibiotics are β -lactams, glycopeptides, polyketides or polypeptide antibiotics. There are also some classes of antibiotics that are prescribed less often as they are used as reserve if other antibiotics fail. This is due to the big problem of resistance that arises with antibiotics. Bacteria are able to develop mechanisms to avoid the knock-out by antibiotics. There are different forms of resistance, for example the modification of the target, *i.e.* the antibiotic does not recognize it anymore, or modification of the antibiotic itself such that it cannot act anymore, or posttranslational modification of the target protein, *i.e.* the strength of the binding of the antibiotic is reduced. Some bacteria are also able to produce efflux proteins that pump the antibiotic out of the cell.

The many existing different bacteria constantly develop new strategies to make the antibiotics useless. Bacteria have actually already developed resistance against the reserve antibiotics, therefore, it is of major importance to handle the use of the antibiotics with care and not to use them too casually. It is not just useless but even dangerous to prescribe antibiotics against viral diseases as antibiotics are ineffective

against viruses. Especially in hospitals the rapid spread of resistant bacteria is an enormous problem and the patients cannot get the help they need. Besides a careful usage of antibiotics the persistent development and research for new antibiotics is of considerable importance. Antibiotics with new modes of action or several different modes of action have a big advantage compared to antibiotics with only one mechanism. [61–63]

4.2 Lantibiotics [64–69]

Lantibiotics (*lanthionine containing antibiotics*) are small antimicrobial peptides (< 5 kDa) [66,67] that are produced by Gram positive bacteria and the antibacterial activity of which is directed towards other Gram positive bacteria. They are ribosomally synthesized and posttranslationally modified by enzymes. They contain unusual amino acids such as lanthionine (Lan) or methyllanthionine (MeLan) (Fig. 4.1) or dehydrated amino acids.

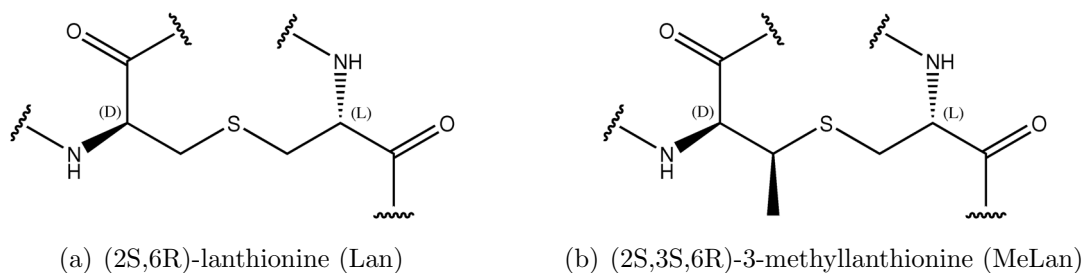
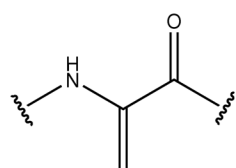
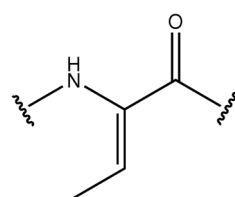


Figure 4.1: Lanthionine and methyllanthionine.

The peptide sequence which is encoded by the gene is called prelantibiotic. It consists of an N-terminal leader sequence and a C-terminal prolantibiotic which undergoes several modifications. The amino acids serine and cysteine are precursors for lanthionine while threonine and cysteine are precursors for methyllanthionine. Serine and threonine are dehydrated enzymatically to form the α, β -unsaturated amino acids 2,3-didehydroalanine (Dha) and 2,3-didehydrobutyryne (Dhb) (Fig. 4.2), respectively.



(a) 2,3-didehydroalanine (Dha)



(b) (Z)-2,3-didehydrobutyryne (Dhb)

Figure 4.2: The α, β -unsaturated amino acids Dha and Dhb.

Those unsaturated amino acids can be modified further when neighboring cysteines add with their thiol groups stereospecifically to the double bond of those unsaturated amino acids to form monosulfide bonds which introduce intramolecular rings to the peptide. The resulting thioether amino acids Lan and MeLan are in meso (DL) configuration which means that a D-alanine residue is introduced from the L-serine (see Fig. 4.3). Lanthionine was first isolated by Horn *et al.* in 1941 when they treated wool with sodium carbonate^[70]. They chose the name lanthionine because it was isolated from wool and contains sulfur.

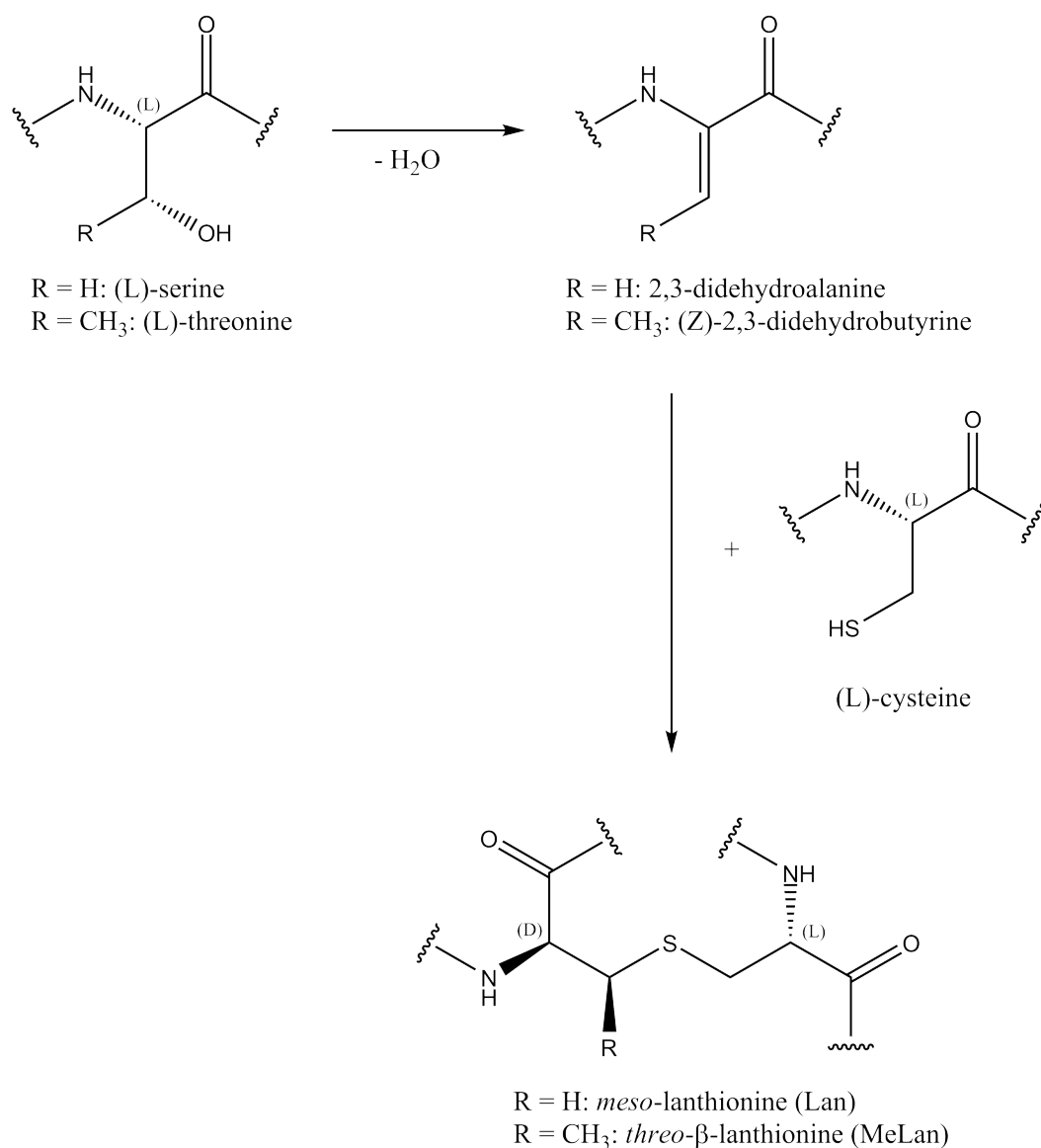


Figure 4.3: Formation of the thioether amino acids Lan and MeLan from Ser and Thr, respectively.

After the modification process the leader sequence is removed by a protease and the biologically active peptide is secreted. Interestingly, Ser and Thr are also found in the leader peptide but these residues are not modified, whereas Cys is only found in the prolantibiotic [65,67,69]. The function of the leader is not clarified yet, but several roles are suggested: protection of the producing strain from the antimicrobial activity by keeping the peptide inactive, assisting in transport of the modifying enzymes towards their target and acting as a recognition sequence for the biosynthetic enzymes.

About 50 lantibiotics are known to date [69]. They are divided into three different classes depending on their shape and biosynthetic pathway. Class I lantibiotics, such

as nisin and subtilin, are modified by two enzymes, which are called LanB and LanC. LanB is responsible for the dehydration of Thr and Ser, whereas the role of LanC is the thioether formation. Class II lantibiotics only have one modifying enzyme called LanM. Members of this class are *e.g.* cinnamycin and mersacidin. Class I lantibiotics are more linear than class II lantibiotics. Class III lantibiotics have no significant antibiotic activity but have other modes of action. ^[71]

One advantage of the antibiotic activity of lantibiotics is the diversity in their effectiveness. The mode of action of lantibiotics is permeabilization of the cell membrane by pore formation on one hand and binding to lipid II, a biosynthetic intermediate responsible for the formation and stabilization of the cell wall, on the other hand. By binding to lipid II lantibiotics inhibit the cross-linking of lipid II which disturbs the formation of a stable cell wall.

One of the best studied lantibiotics is nisin, produced by the Gram positive bacterium *Lactococcus lactis* and used as food-preservative. Although it is already used for more than 40 years now, no resistances against it have been detected, yet ^[69], which might be due to the diversity of the mode of action of this lantibiotic.

Some other important lantibiotics are epidermin, lacticin 481, cinnamycin or mersacidin.

Labyrinthopeptin A2 (from the greek word labyrinthos) is a lantibiotic which is produced by the Gram positive strain *Actinomadura namibiensis* ^[72]. It consists of 18 amino acids and is strongly hydrophobic. It has two lanthionines and a disulfide bridge between two cysteines which form five rings (A, B, A', B', C) (see Fig. 4.4). It also has an unchanged threonine which was not modified enzymatically. The leader sequence was found to consist of 20 amino acids. A2 has been characterized as a class III lantibiotic because of its structural features and as it shows only slight antibacterial but more distinct analgetic activity. ^[73,74]

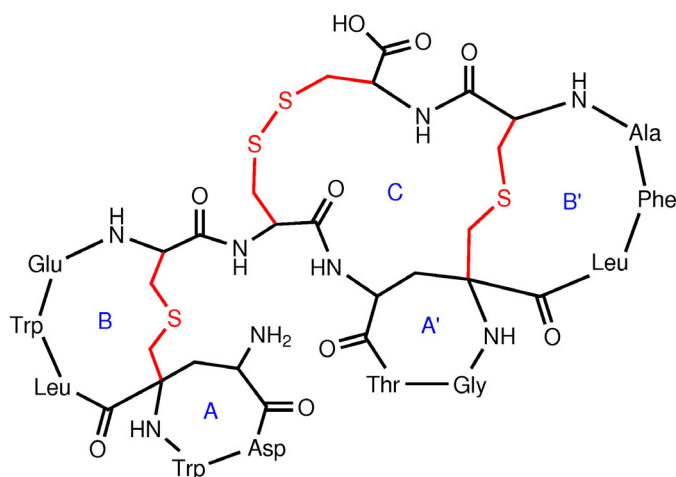


Figure 4.4: Structure of labyrinthopeptin A2 with the denotation of the rings.

4.3 Crystallization

4.3.1 Basics and principles in crystallization of macromolecules

Crystallization of a protein depends on many different factors. Variation of pH, concentration and kind of buffer, concentration and kind of salt, temperature, dielectric constant, protein concentration and many more can decide on whether crystals grow or not. A phase diagram as shown in Fig. 4.5 can visualize the influence of *e.g.* the salt concentration and the protein concentration on formation of crystals. In this diagram, the concentration of salt and protein are varied, while all other parameters are kept constant. The area of undersaturation is separated by the solubility curve from the area of supersaturation. If the concentration of the protein is so low that the solution is undersaturated, the protein cannot crystallize. Supersaturation is essential for the formation of crystals. The supersaturation zone can be subdivided into three zones. In the precipitation zone, the protein will precipitate as an amorphous solid, but not in crystalline form. In the nucleation zone, the protein nucleates in form of many small crystals. It can happen that saturation is reached but no crystal big enough for an X-ray diffraction measurement has formed. One has then the possibility to put a small crystal in a protein solution with a concentration that corresponds to conditions of the metastable zone, as in this zone crystallization may only happen when *e.g.* seeds of small crystals are introduced to this solution. Crystals can grow under these conditions slow enough to form few big crystals suitable for X-ray diffraction without nucleation of new crystals. ^[75]

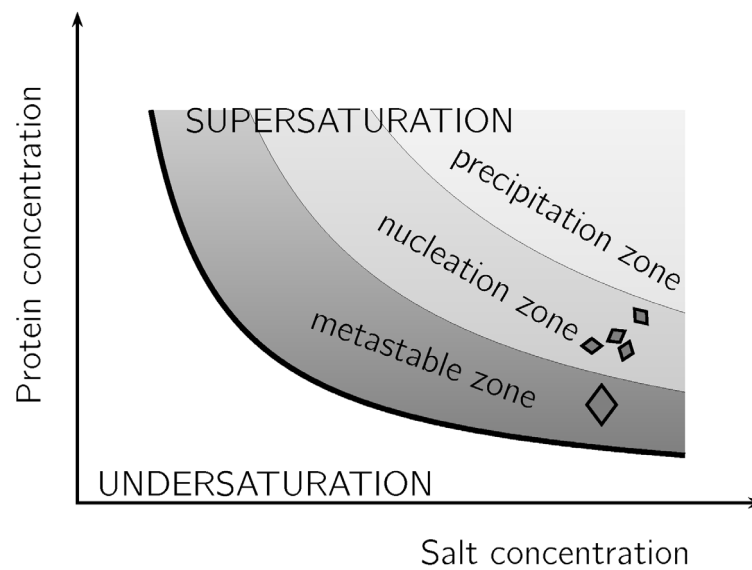


Figure 4.5: Schematic phase diagram for the crystallization of a protein.

4.3.1.1 Hanging drop vapor diffusion method

By the hanging drop vapor diffusion method the peptide solution is mixed in a drop on a siliconized cover slide with the reagent solution, which consists of salt, buffer, precipitant or other additives. The drop is equilibrated against a solution containing the reagents. As the concentration of the reagents in the drop is much lower than in the solution an equilibrium is reached by evaporation of water which moves from the drop into the solution (see Fig. 4.6). This leads to a supersaturation of the protein in the drop, which hopefully results in crystallization of the protein. This method can only be used if the drop is not too big as otherwise it would drop into the reagent solution instead of being equilibrated.

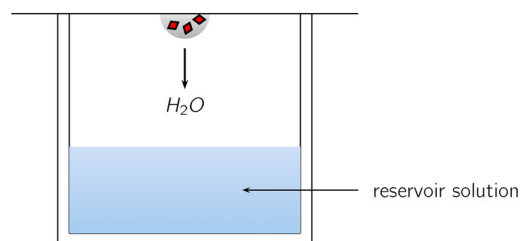


Figure 4.6: Hanging drop vapor diffusion crystallization.

4.3.1.2 Sitting drop vapor diffusion method

The sitting drop vapor diffusion method is similar to the hanging drop vapor diffusion method. Here again, a solution containing the peptide and the reagent solution are

equilibrated against each other (Fig. 4.7). As the drop is sitting on a bridge this method is suitable especially for large drops and drops containing reagents with a low surface tension.

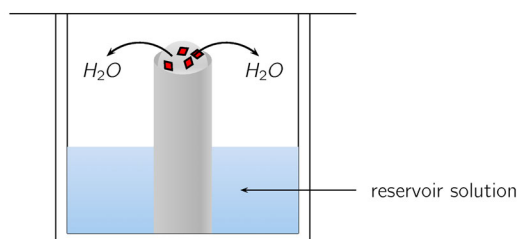


Figure 4.7: Sitting drop vapor diffusion crystallization.

4.3.1.3 Cryocrystallography

When the measurement is performed at room temperature there are several problems that can occur. One is that radiation damage can destroy the crystal before the measurement can be finished successfully. Radiation damage occurs by primary interaction between the beam and the molecules in the crystal. The energy of the beam produces heat and thus increases the thermal vibration of the molecules, and this energy is also sufficient to break certain bonds in the molecules. Radicals that are produced in this way can then further diffuse through the crystal and cause even more damage. ^[76,77] When the crystal is cooled down to 100 K the problem of radiation damage is less distinct. But as proteins often contain a lot of water care must be taken when the crystal is cooled down. It must be avoided that water freezes to ice and thus cracks the crystal. This can be achieved by soaking the crystal in an anti-freeze agent (cryoprotectant) and flash-cooling it down to 100 K. ^[78,79] The anti-freeze agent prevents the water from freezing to ordered ice but water forms a vitreous glass, instead. When the molecules in the protein crystal are frozen vibration is reduced which allows the crystal to diffract to higher resolution. Also the degree of disorder (at least of dynamic disorder) is reduced at lower temperatures.

4.3.2 Crystallization conditions for A2

Crystallization conditions for A2 were found by screening with the [Jena Bioscience](#) ^[80] JBScreen Classic 8 when the conditions of the two neighboring wells D4 (60 % w/v ethanol, 1.5 % w/v PEG 6000, 0.05 M sodium acetate) and D5 (60 % w/v ethanol, 0.10 M sodium chloride) were accidentally mixed. As the concentration of the resulting solution was not exactly known for this reason, a 24 well plate was set up with

varying conditions for ethanol, PEG 6000, sodium acetate and sodium chloride. The composition which yielded the best crystals was a reservoir mixture of 60 % ethanol, 0.75 % PEG 6000, 0.025 M sodium acetate and 0.05 M sodium chloride. For the crystallization the peptide was dissolved in 0.02 M TRIS of pH 8.2 with a concentration of 7 mg/ml. For the drop 2 μ l of peptide were mixed with 2 μ l of the reservoir solution. Crystals grew within one week to a size of $0.125 \times 0.075 \times 0.025$ mm³ at room temperature by using the sitting drop vapor diffusion method. A crystal of A2 is shown in Fig. 4.8. The crystals were soaked for a few seconds in a solution containing 15 % 2,3-butanediol as a cryoprotectant. Subsequently, the crystal was flash frozen in liquid nitrogen and transferred quickly to the diffractometer.



Figure 4.8: Crystal of labyrinthopeptin A2.

X-ray data were collected on a Bruker three circle diffractometer equipped with a rotating anode with mirror-monochromated Cu-K α radiation ($\lambda = 1.54178$ Å). Intensities were measured with a SMART 6000 detector to a resolution of 1 Å.

4.4 Data processing and structure

Integration, absorption correction and scaling of the data was performed with the programs SAINT [81] and SADABS [82]. The space group was determined with the program XPREP [83]. The structure was solved by direct methods with the program SHELXD [84]. The initial assignment of the non-hydrogen atoms of A2 was done by hand with the program XP [85]. All subsequent comparison of the model (σ_A -weighted $2F_{obs} - F_{calc}$ map [86]) with the measured density and the difference density (σ_A -weighted $F_{obs} - F_{calc}$ map) was done with COOT [87]. A least-squares refinement against F^2 for all data was performed with the program SHELXL [10]. Restraints for the distances and angles of the residues were applied according to Engh and Huber [88]. SHELXWAT [11] was used to place oxygen atoms to non-interpreted maxima in the electron density map. As after this procedure there were still some

significant residual density peaks present, which were in a possible hydrogen bond distance to other water molecules, those residual density peaks were converted to oxygen atoms by hand.

A sodium atom is coordinated (in an octahedral geometry) by oxygen atoms from the peptide as well as from oxygen atoms from water molecules. The asymmetric unit also contains one acetate molecule. The peptide has a solvent (water) content of about 20 %. The peptide crystallizes in the orthorhombic space group $P2_12_12$ with the cell constants $a = 41.136 \text{ \AA}$, $b = 12.885 \text{ \AA}$, $c = 25.590 \text{ \AA}$. There is one molecule of A2 in the asymmetric unit. Crystallographic details are given in Tables 4.1 and 4.2.

molecular formula	$C_{86}H_{109.5}N_{20}O_{49.5}S_4Na$
space group	$P2_12_12$
cell parameters (\AA)	$a = 41.136$ $b = 12.885$ $c = 25.590$
resolution (\AA)	1.00 (1.10–1.00)
measured reflections	268732 (43178)
independent reflections	14199
completeness (%)	100 (100)
redundancy	18.87 (12.15)
mean $I/\sigma(I)$	27.57 (3.22)
R_{int} (%)	5.95 (57.24)

Table 4.1: Data collection statistics for labyrinthopeptin A2; values in parentheses denote the outer resolution shell.

Final R_1 ($F > 4\sigma(F)$) (%)	11.39
Final R_1 (all data) (%)	13.25
Final wR_2 (%)	29.00
$Goof$	1.255
Flack x parameter ^[89]	0.0251(326)
Number of peptide atoms	163
Number of solvent atoms	43
Number of metal ions	1
solvent content (%)	18.7

Table 4.2: Refinement statistics for labyrinthopeptin A2.

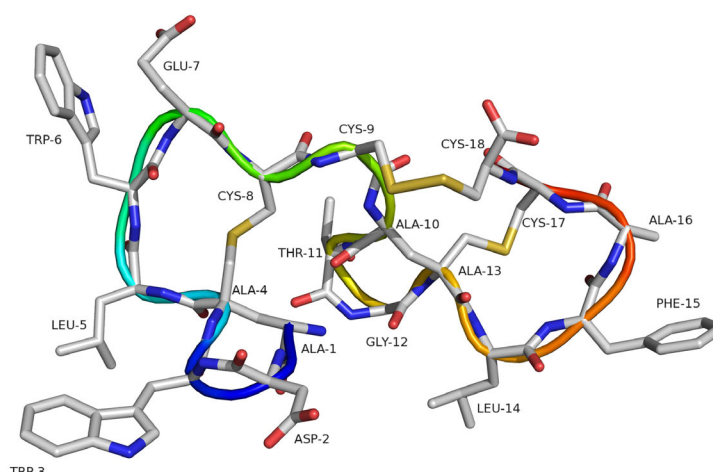


Figure 4.9: Structure of A2; the rainbow colored chain traces the C_{α} atoms from the N- towards the C-terminus.

A2 consists of 18 amino acids with the sequence Ala¹-Asp²-Trp³-Ala⁴-Leu⁵-Trp⁶-Glu⁷-Cys⁸-Cys⁹-Ala¹⁰-Thr¹¹-Gly¹²-Ala¹³-Leu¹⁴-Phe¹⁵-Ala¹⁶-Cys¹⁷-Cys¹⁸ (see Figs. 4.9 and 4.10).

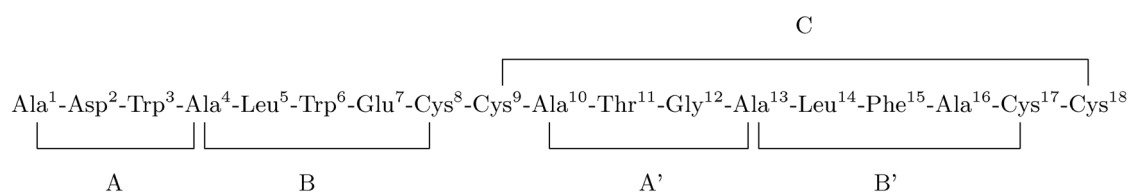


Figure 4.10: Amino acid sequence of the lantibiotic A2 with intramolecular connections and denotation of the rings.

Cys⁹ and Cys¹⁸ are connected by a disulfide bridge. Ala⁴ and Cys⁸ as well as Ala¹³ and Cys¹⁷ each form the amino acid lantionine and thus are connected by a mono sulfide bridge. Ala⁴ and Ala¹³ are amino acids with quaternary C_{α} atoms that are coordinated not only three-fold (which is normally typical for peptides and proteins) but even four-fold. This leads to a novel structural motif of two directly connected rings and to the existence of five rings in the molecule which are labeled A, B, A', B' and C. The rings A and A' consist of the amino acids Ala¹-Asp²-Trp³-Ala⁴ (ring A) and Ala¹⁰-Thr¹¹-Gly¹²-Ala¹³ (ring A') and the ring closure is due to the quaternary amino acids Ala⁴ and Ala¹³, respectively. Rings B and B' are closed by the amino acid lantionine and are built up by Ala⁴-Leu⁵-Trp⁶-Glu⁷-Cys⁸ (ring B) and Ala¹³-Leu¹⁴-Phe¹⁵-Ala¹⁶-Cys¹⁷ (ring B'). Ring C is formed by the linkage of the disulfide bridge between Cys⁹ and Cys¹⁸ and consists of Cys⁹-Ala¹⁰-Thr¹¹-Gly¹²-Ala¹³-Leu¹⁴-Phe¹⁵-Ala¹⁶-Cys¹⁷-Cys¹⁸. As the small rings A and A' are

built up from only 4 amino acids each (11 atoms) this leads to the rare motif of a *cis*-peptide bond between Asp2 and Trp3 in the A-ring and Thr11 and Gly12 in the A'-ring, respectively. The rings A and B, and A' and B', respectively, are connected via the two amino acids that contain a quaternary C_α, Ala4 and Ala13, respectively.

A second position for the atoms CB, OG1 and CG2 in Thr11 could be determined and was refined to an occupation factor of 50%. The positions of the sulfur atoms in Cys9 and Cys18 are disordered, too, and build an alternative orientation for the disulfide bridge (see Fig. 4.11).

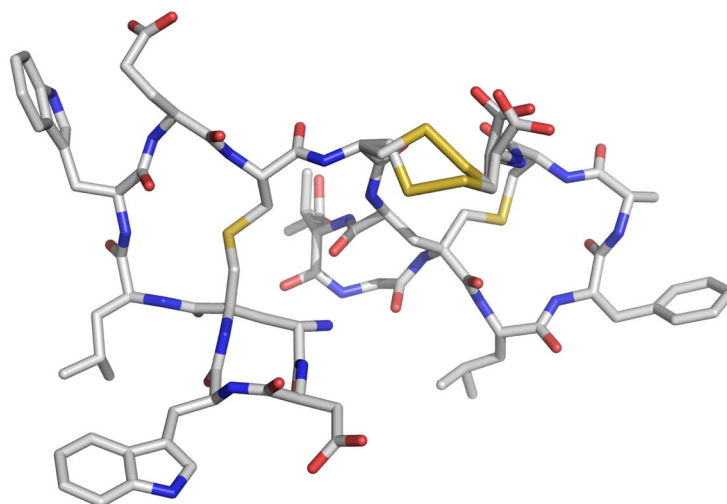


Figure 4.11: Structure of labyrinthopeptin A2 with refined disordered regions.

The electron density in the region of the residues from Leu14 to Ala16 is quite flexible, but was not refined as disordered because no distinct two positions for each atom could be determined. This region seems to be dynamically disordered.

4.4.1 Nomenclature of the atoms in the residues in A2

The atoms in A2 were labeled according to the IUPAC nomenclature on amino acids^[90]. An overview on the nomenclature that is used throughout the text is given in Fig. 4.12.

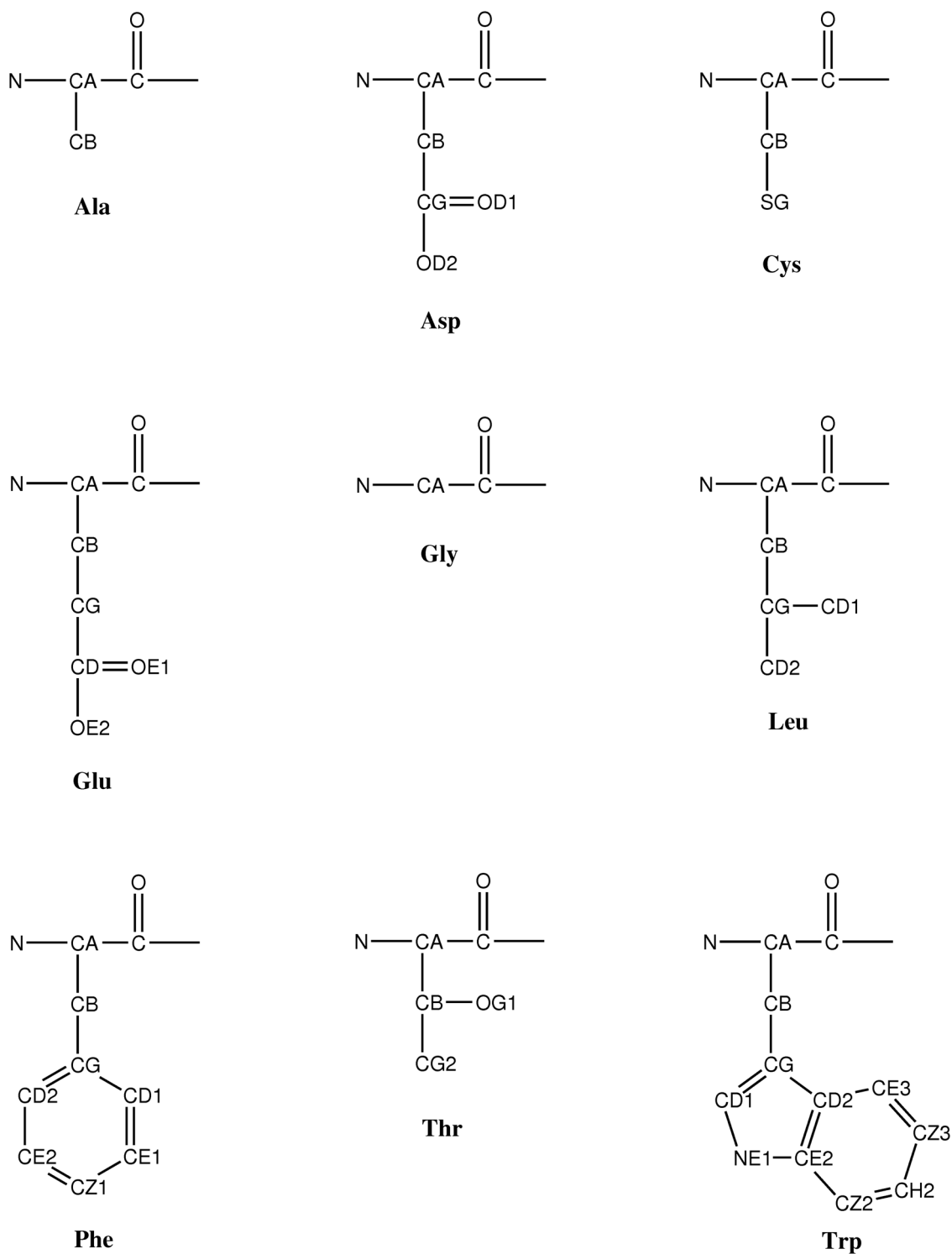


Figure 4.12: Amino acids in labyrinthopeptin A2 with respective nomenclature of the atoms; in atom names consisting of two letters the first letter denotes the atom type and the second letter is the abbreviation for the respective greek letter (A = α , B = β , G = γ , D = δ , E = ϵ , Z = ζ , H = η).

4.4.2 Ramachandran plot

For a structure validation of A2 a Ramachandran plot^[91] was created with the program PROCHECK^[92]. This plot provides a consistency check for the dihedral angles φ and ψ . It was shown^[93] that there are only a few sets of φ and ψ which are allowed in proteins for all amino acids except for glycine due to steric clashes. The Ramachandran plot can be divided into three main regions (A, B and L) which are further subdivided into regions depending on how favored the corresponding angle sets are. The colors indicate the preference for the respective regions. The most favored regions A, B and L are in dark red, the additional allowed regions a, b and l are in dark yellow and the generously allowed regions -a, -b and -l are in light yellow. Amino acids in the A-regions (A, a and -a) are mainly found in right handed α -helices, those in the B-regions (B, b and -b) in β -strands and the amino acids in the L-regions (L, l and -l) mainly in left handed α -helices. All dihedral angles in A2 lie within the allowed regions (see Fig. 4.13).

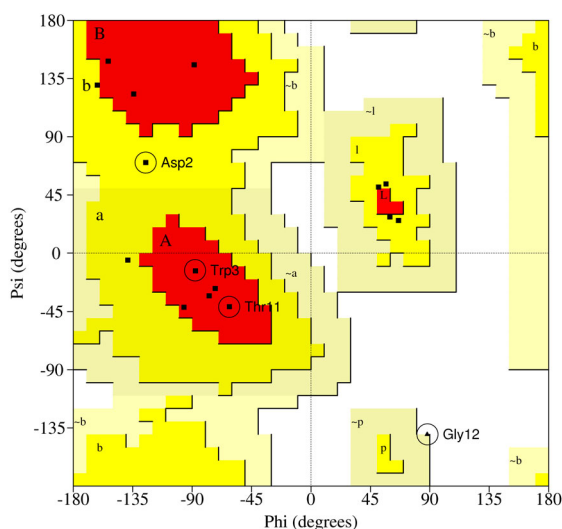


Figure 4.13: Ramachandran plot for labyrinthopeptin A2; amino acids involved in *cis* bonds are highlighted.

4.4.3 Cis peptide bonds

The peptide bond between two amino acids in proteins or peptides is formed when the carboxyl group of one amino acid and the amino group of the other amino acid react with each other under cleavage of water. Due to its partial double bond character the peptide bond is planar and there is a rotational barrier between the *trans* ($\omega = 180^\circ$) and *cis* ($\omega = 0^\circ$) conformation^[90] (Fig. 4.14) of about

13–20 kcal/mol^[94–96] (for an overview on the nomenclature of the dihedral angles see Fig. 4.15).



Figure 4.14: *Trans* and *cis* peptide bonds.

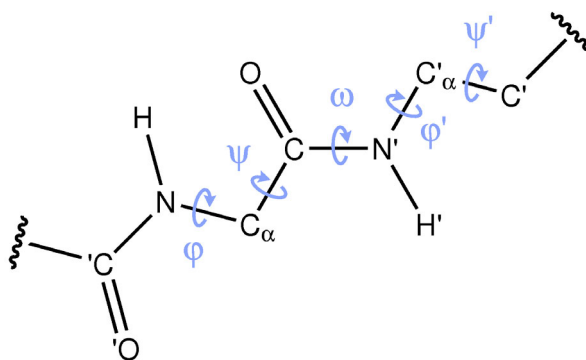


Figure 4.15: Torsion angles in a peptide bond; φ : 'C–N–C_α–C, ψ : N–C_α–C–N', ω : C_α–C–N'–C'_α; the prime before and after an atom name means that the corresponding atom is part of the preceding and succeeding residue relative to the discussed amino acid, respectively.

Due to steric repulsion of the two C_α carbon atoms the *cis* conformation is energetically less favored. An energy difference between *cis* and *trans* of 0.5 kcal/mol^[94,96,97] was found for Xaa–Pro (*i.e.* any amino acid bonded to proline) peptide bonds and of 2.5–4 kcal/mol^[94,96,97] for Xaa–nonPro bonds (Fig. 4.16). The reason for the lower energy difference in *cis* Xaa–Pro peptide bonds is that the C_δ atom in the proline also interacts with the C_α atom in the other amino acid. From the energy differences in *cis* and *trans* conformations it can be calculated that about 30 % of all Xaa–Pro and 1.5 % of all Xaa–nonPro peptide bonds should be in *cis* conformation. The values found (2.5–4 % and 0.03–0.05 %) are much smaller. One reason for this discrepancy could be that at low resolution it can easily happen that a *cis* conformation is overlooked and refined as a *trans* bond. This is also in accordance with the resolution dependence of the frequency of detected *cis* peptide bonds^[97,98].

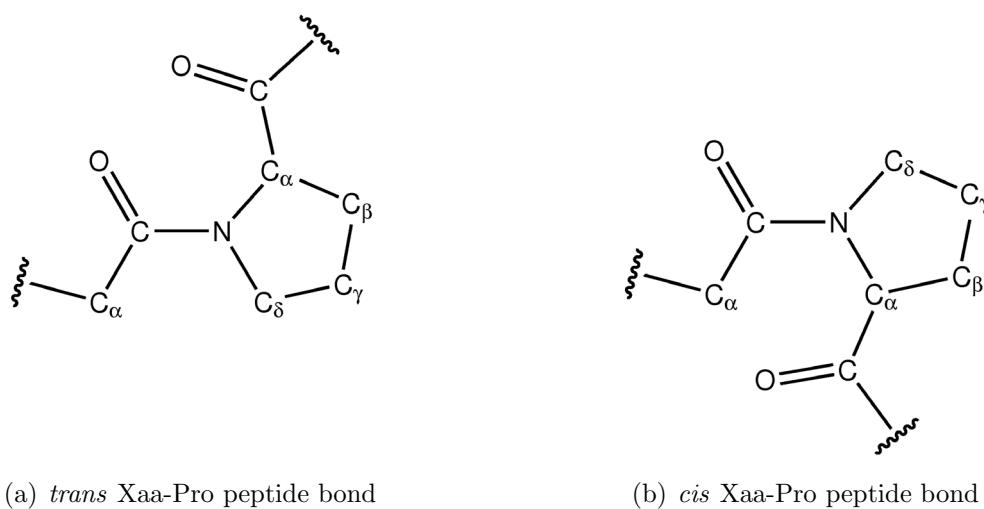


Figure 4.16: *Trans* and *cis* Xaa-Pro peptide bonds.

Cis peptide bonds occur very rarely, only about 0.3% of all peptide bonds are *cis* and more than 80% among them are Xaa-Pro *cis* bonds [94,96–98]. This makes 0.003–0.005% of all peptide bonds *cis* Xaa–nonPro (amide) peptide bonds and 4.7–6.5% *cis* Xaa–Pro (imide) bonds. In A2 two of the 17 peptide bonds are in *cis* conformation, Asp2–Trp3 ($\omega = 14.419^\circ$) and Thr11–Gly12 ($\omega = 1.293^\circ$) (Table 4.3 and Fig. 4.17).

	ω ($^{\circ}$)
Ala1-Asp2	172.0
Asp2-Trp3	14.4
Trp3-Ala4	173.5
Ala4-Leu5	172.3
Leu5-Trp6	-178.1
Trp6-Glu7	176.6
Glu7-Cys8	-179.1
Cys8-Cys9	164.7
Cys9-Ala10	178.0
Ala10-Thr11	-172.1
Thr11-Gly12	1.3
Gly12-Ala13	169.2
Ala13-Leu14	-173.0
Leu14-Phe15	-178.1
Phe15-Ala16	174.7
Ala16-Cys17	178.1
Cys17-Cys18	-177.6/-168.9

Table 4.3: ω torsion angles in labyrinthopeptin A2; the two values for the ω angle between Cys17 and Cys18 result from the disorder in Cys18.

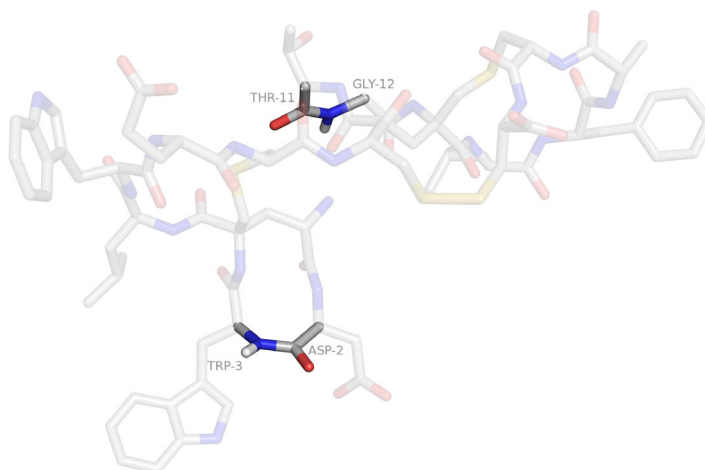


Figure 4.17: *Cis* peptide bonds between Asp2 and Trp3 and between Thr11 and Gly12 in labyrinthopeptin A2.

Jabs *et al.*^[98] found that *cis* peptide bonds in small rings occur more frequently than in acyclic systems. They searched the CSD (Cambridge Structural Database^[99]) and found 269 out of 527 (> 50 %) peptide bonds in cyclic molecules to be in *cis* conformation, preferably in small rings with not more than 12 atoms in the ring.

In small rings like the A and A' rings in A2, which consist of 11 atoms each, the *cis* bond is formed for sterical reasons. If the peptide bonds Asp2–Trp3 and Thr11–Gly12 in A2 were in *trans* conformation the carbonyl oxygen atoms or the amine hydrogen atoms would be directed towards the center of the ring which is not possible for steric reasons. Jabs *et al.* also found all residues N-terminal and most residues C-terminal to the *cis* peptide bond to be in the B-region of the Ramachandran plot. In contrast to these findings, in A2 the N-terminal residue Trp3 is in the A-region whereas Gly12 lies in a special region as is expected for glycine, anyway. The C-terminal residue Thr11 is also in the A-region whereas Asp2 indeed lies in the B-region. According to Jabs *et al.* the majority of the ψ_1/φ_2 angles lie in the region of $(\psi_1, \varphi_2) = (+110.3^\circ, -136.2^\circ)$ and $(\psi_1, \varphi_2) = (+158.3^\circ, -102.0^\circ)$. This corresponds to the angle pairs of $(+80^\circ, -150^\circ)$ and $(+150^\circ, -80^\circ)$, which have been calculated in a conformational study to be the only possible angles in a *cis* peptide bond. In contrast to these findings the (ψ_1, φ_2) angles in A2 have the values $(-70.1^\circ, -87.4^\circ)$ for Asp2–Trp3 and $(-41.2^\circ, 88.0^\circ)$ for Thr11–Gly12 (see Table 4.4). In the Ramachandran plot the positions of the φ and ψ angles for Asp2, Trp3, Thr11 and Gly12 are marked with a circle (see Fig. 4.13).

Jabs *et al.* reported that none of their investigated structures showed a positive φ angle at the C-terminus of the *cis* peptide bond. Both of the C-terminal φ angles in A2 are negative, indeed $(\varphi(\text{Asp2}) = -125.1^\circ, \varphi(\text{Thr11}) = -61.8^\circ)$; see Table 4.4).

	φ (°)	ψ (°)
Ala1		126.3
Asp2	-125.1	70.1
Trp3	-87.5	-13.5
Ala4	56.8	53.5
Leu5	-96.2	-41.7
Trp6	-153.6	148.4
Glu7	51.2	51.1
Cys8	-88.4	145.6
Cys9	-72.5	-27.2
Ala10	-134.3	123.0
Thr11	-61.8	-41.2
Gly12	88.0	-140.1
Ala13	66.3	25.2
Leu14	-77.1	-32.9
Phe15	-161.8	129.9
Ala16	59.8	28.0
Cys17	-138.7	-5.4/-7.8
Cys18	-101.2/-78.6	

Table 4.4: φ and ψ torsion angles in labyrinthopeptin A2; the two values for the φ and ψ angles between Cys17 and Cys18 result from the disorder in Cys18.

In *cis* peptide bonds the C_α -C-N' as well as the C-N'-C' $_\alpha$ angles are found to be widened compared to non *cis* peptide bonds^[98]. The reason for this finding is a steric repulsion of the two neighboring C_α atoms. This holds true for A2 as the CA_2-C_2-N_3 and CA_11-C_11-N_12 angles are 121.3° and 126.2°, respectively. Except for the C-terminal disordered residue Cys18 those angles are much larger than all the other C_α -C-N' angles in A2. Also the angles between C_2-N_3-CA_3 (127.3°) and C_11-N_12-CA_12 (128.7°) are widened compared to the corresponding other angles in A2 (except for the respective angle between Cys17 and Cys18, again; see Table 4.5).

	$C_{\alpha}-C-N'$ ($^{\circ}$)	$C-N'-C'_{\alpha}$ ($^{\circ}$)
Ala1-Asp2	115.1	126.0
Asp2-Trp3	121.3	127.3
Trp3-Ala4	120.0	121.0
Ala4-Leu5	117.0	122.1
Leu5-Trp6	118.7	122.6
Trp6-Glu7	116.6	122.0
Glu7-Cys8	114.2	121.4
Cys8-Cys9	116.1	118.6
Cys9-Ala10	117.0	121.7
Ala10-Thr11	116.8	115.7
Thr11-Gly12	126.2	128.7
Gly12-Ala13	114.5	121.7
Ala13-Leu14	117.2	119.3
Leu14-Phe15	117.1	124.8
Phe15-Ala16	117.0	120.8
Ala16-Cys17	118.7	123.6
Cys17-Cys18	123.2/117.8	127.5/109.4

Table 4.5: $C_{\alpha}-C-N'$ and $C-N'-C'_{\alpha}$ angles in labyrinthopeptin A2; the respective angles for the amino acids involved in the *cis* peptide bonds (Asp2/Trp3 and Thr11/Gly12) are larger than the corresponding angles in *trans* peptide bonds; the two values for the angles between Cys17 and Cys18 result from the disorder in Cys18.

4.4.4 Quaternary D amino acids

Naturally occurring amino acids are normally in the L conformation (see Fig. 4.18). D amino acids often originate from L amino acids which were posttranslationally modified to D amino acids.



Figure 4.18: L and D isomers of amino acids.

In the case of A2 the two D amino acids result from the formation of the lanthionine from an L-serine and an L-cysteine (see Fig. 4.3). During the formation of lanthionine a water molecule is cleaved off from the serine which leads to a dou-

ble bond in the resulting amino acid. The attack of the cysteine is such that the conformation of the amino acid is changed into the D form. The C_α atoms in A2 that come from the L-serine are the two D amino acids Ala4 and Ala13. These two amino acids are not only in the rare D form but their C_α atoms are also quaternary coordinated. Only few quaternary amino acids are structurally determined. Bunkóczy *et al.* report on crystal structures of cephaibols that possess amino acids with quaternary C_α atoms with two methyl residues.^[100] But in contrast to Ala4 and Ala13 in A2 these methyl residues do not form bridges to other atoms. In A2 the quaternary C_α atoms connect the A ring with the B ring by linking CA_4 with CB_1, and the A' ring with the B' ring by linking CA_13 with CB_10, respectively. By a search in the CSD (version 5.29) no similar structural motif was found.

4.4.5 Comparison of the rings in A2

4.4.5.1 Overlay of the rings A and A'

The residues Ala1–Ala4 were aligned with the residues Ala10–Ala13 to analyze the similarities between the A and the A' ring. For a comparison of the structural features of A and A' a pair-fitting was performed with the program PyMOL^[101]. An alignment of CA_1:CA_10, CB_1:CB_10, N_4:N_13, CA_4:CA_13 shows (see Fig. 4.19) that the rings A and A' deviate mainly in the residues with the *cis* peptide bonds (Asp2/Trp3 and Thr11/Gly12).

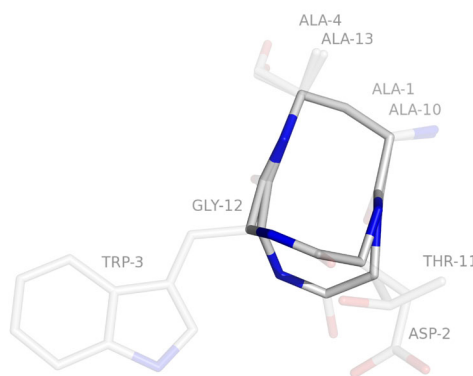


Figure 4.19: Overlay of ring A with A' in labyrinthoheptin A2.

The residues Thr11 and Gly12 are mainly oriented above the plane in which the ring atoms lie, whereas the residues Asp2 and Trp3 are oriented below the plane. The difference in the involved angles between rings A and A' is shown in Table 4.6. This might be due to the bigger flexibility in ring A' because of Gly12.

	ring A (°)		ring A' (°)
CB_1-CA_1-C_1	112.5	CB_10-CA_10-C_10	111.8
CA_1-C_1-N_2	115.1	CA_10-C_10-N_11	116.7
C_1-N_2-CA_2	126.0	C_10-N_11-CA_11	115.7
N_2-CA_2-C_2	106.8	N_11-CA_11-C_11	113.2
CA_2-C_2-N_3	121.3	CA_11-C_11-N_12	126.2
C_2-N_3-CA_3	127.3	C_11-N_12-CA_12	128.7
N_3-CA_3-C_3	116.0	N_12-CA_12-C_12	108.3
CA_3-C_3-N_4	120.0	CA_12-C_12-N_13	114.5
C_3-N_4-CA_4	121.0	C_12-N_13-CA_13	121.7
N_4-CA_4-C_1	113.0	N_13-CA_13-C_10	111.2
CA_4-CB_1-CA_1	116.5	CA_13-CB_10-CA_10	115.2

Table 4.6: Angles in the A and A' rings in labyrinthopeptin A2.

For an alignment of the complete ring without the side chain atoms (CA_1:CA_10, C_1:C_10, CB_1:CB_10, N_2:N_11, CA_2:CA_11, C_2:C_11, N_3:N_12, CA_3:CA_12, C_3:C_12, N_4:N_13, CA_4:CA_13) an RMSD of 0.690 Å was calculated.

4.4.5.2 Overlay of the rings B and B'

The residues Ala4 to Cys8 in ring B were superimposed with the corresponding residues Ala13 to Cys17 in ring B' to compare similarities and differences in both rings.

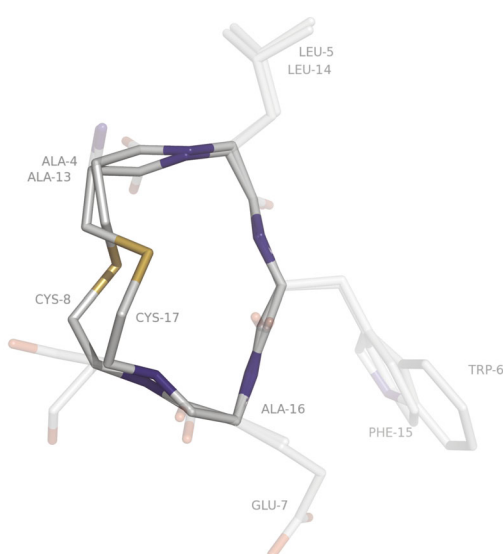


Figure 4.20: Overlay of ring B with B' in labyrinthopeptin A2.

For the investigation of the angles and the orientation of respective atoms in the two rings the alignment was done for the atoms C_5:C_14, N_6:N_15, CA_6:CA_15, C_6:C_15, N_7:N_16, CA_7:CA_16. The biggest differences in the angles are in the atoms of the two lanthionines, especially the two sulfur atoms (Ala4/Cys8 and Ala13/Cys17; see Fig. 4.20). The values for the angles are shown in Table 4.7.

	ring B (°)		ring B' (°)
CB_4-CA_4-C_4	108.2	CB_13-CA_13-C_13	109.3
CA_4-C_4-N_5	117.0	CA_13-C_13-N_14	117.2
C_4-N_5-CA_5	122.1	C_13-N_14-CA_14	119.3
N_5-CA_5-C_5	111.9	N_14-CA_14-C_14	117.2
CA_5-C_5-N_6	118.7	CA_14-C_14-N_15	117.1
C_5-N_6-CA_6	122.6	C_14-N_15-CA_15	124.8
N_6-CA_6-C_6	106.5	N_15-CA_15-C_15	110.6
CA_6-C_6-N_7	116.6	CA_15-C_15-N_16	117.0
C_6-N_7-CA_7	122.0	C_15-N_16-CA_16	120.8
N_7-CA_7-C_7	109.2	N_16-CA_16-C_16	113.7
CA_7-C_7-N_8	114.2	CA_16-C_16-N_17	118.7
C_7-N_8-CA_8	121.4	C_16-N_17-CA_17	123.6
N_8-CA_8-CB_8	111.1	N_17-CA_17-CB_17	109.3
CA_8-CB_8-SG_8	114.3	CA_17-CB_17-SG_17	114.5
CB_8-SG_8-CB_4	100.7	CB_17-SG_17-CB_13	99.7
SG_8-CB_4-CA_4	111.9	SG_17-CB_13-CA_13	113.7

Table 4.7: Angles in the B and B' rings in labyrinthopeptin A2.

The RMSD value for the alignment of the two rings B and B' without their side chain atoms (CA_4:CA_13, C_4:C_13, CB_4:CB_13, N_5:N_14, CA_5:CA_14, C_5:C_14, N_6:N_15, CA_6:CA_15, C_6:C_15, N_7:N_16, CA_7:CA_16, C_7:C_16, N_8:N_17, CA_8:CA_17, CB_8:CB_17, SG_8:SG_17) was calculated to be 0.781 Å.

4.4.5.3 Overlay of the rings AB and A'B'

For comparison of the relative orientation of the AB ring to the A'B' ring an alignment of the two quaternary C_α atoms and their bonding atoms (CB_1:CB_10, N_4:N_13, CA_4:CA_13, C_4:C_13, CB_4:CB_13) was performed (see Fig. 4.21). As already indicated by the individual alignments of rings A/A' and B/B' they deviate only slightly in their orientation. The angles in the two ring systems between rings A and B, and A' and B' are 87.7° and 90.4°, respectively. They were calculated from the dihedral angles between the planes spanned by N_4/N_13, CA_4/CA_13

and CB_1/CB_10 and the planes of C_4/C_13, CA_4/CA_13 and CB_4/CB_13, respectively.

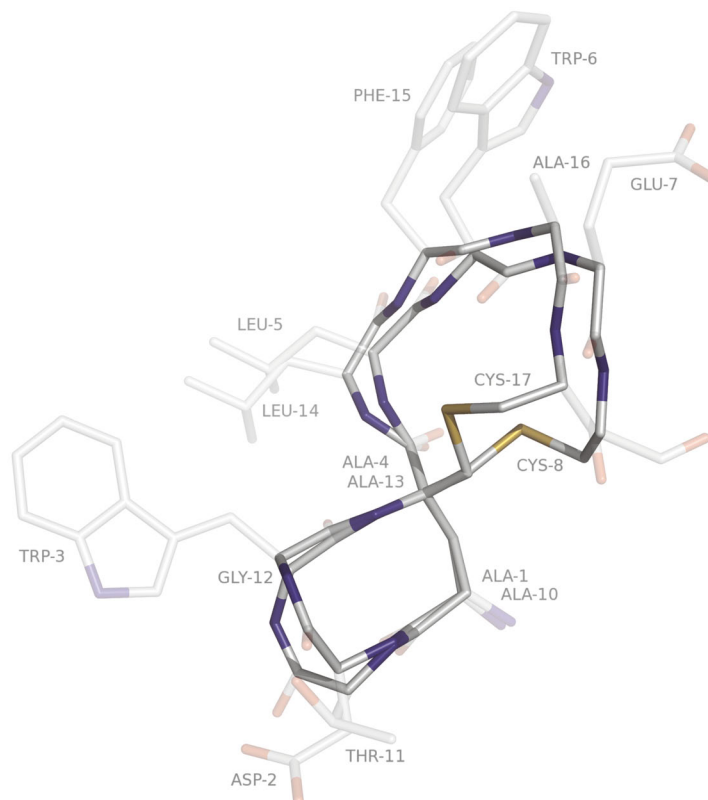


Figure 4.21: Overlay of ring AB with A'B' in labyrinthopeptin A2.

An alignment of the atoms CA_1:CA_10, C_1:C_10, CB_1:CB_10, N_2:N_11, CA_2:CA_11, C_2:C_11, N_3:N_12, CA_3:CA_12, C_3:C_12, N_4:N_13, CA_4:CA_13, C_4:C_13, CB_4:CB_13, N_5:N_14, CA_5:CA_14, C_5:C_14, N_6:N_15, CA_6:CA_15, C_6:C_15, N_7:N_16, CA_7:CA_16, C_7:C_16, N_8:N_17, CA_8:CA_17, CB_8:CB_17, SG_8:SG_17 yielded an RMSD of 0.951 Å.

4.4.6 Surface

The solvent accessible surface^[102] was calculated with the program PyMOL (see Fig. 4.22). The solvent accessible surface is calculated by rolling a solvent molecule of a radius of 1.4 Å (approximating the radius of a water molecule) over the van der Waals surface of the peptide^[103]. The area that the center of the sphere can touch is defined as the solvent accessible surface. For one single molecule of A2 this area was calculated to be 1916.3 Å².

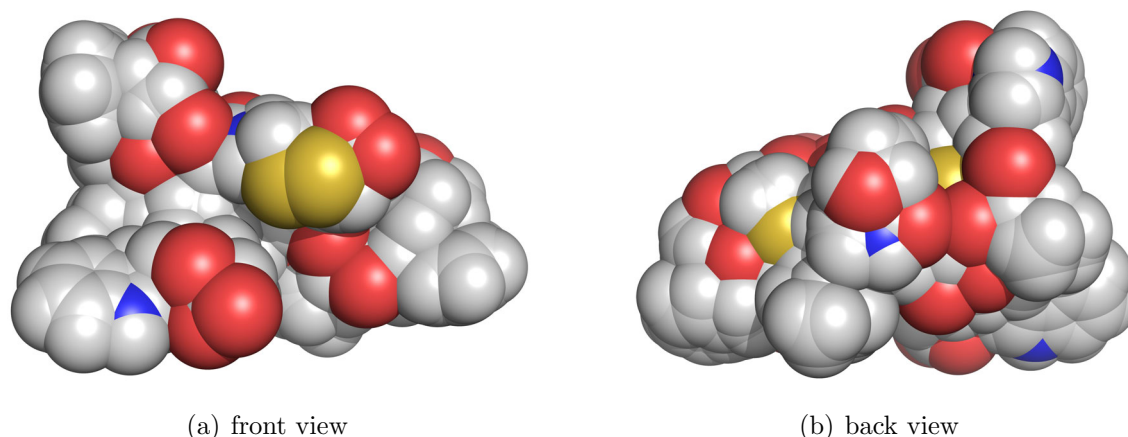


Figure 4.22: Solvent accessible surface of labyrinthopeptin A2.

To investigate whether two molecules of A2 (which are generated by rotating about the twofold axis) come close enough to each other to build a buried surface (the surface which is covered by the close contact of two molecules and which is not accessible by the solvent anymore) two pdb files were produced by rotating one molecule about a twofold axis in both directions. The solvent accessible surfaces of those two “dimers” were calculated with the program PyMOL and resulted in a value of 3463.0 \AA^2 for one dimer and 3103.4 \AA^2 for the other dimer. The buried surface can now be calculated with

$$\text{BS} = \frac{2 \cdot \text{S1M} - \text{S2M}}{2} \quad (4.1)$$

with BS: buried surface, S1M: surface of one molecule, S2M: surface of a dimer. This gave a value of 369.7 \AA^2 for one dimer and 729.3 \AA^2 for the other dimer, which corresponds to 19.3% and 38.1% of the solvent accessible surface of the single molecule, respectively. The first value is maybe too small to suggest a real dimerization of A2, but the other value for the buried surface shows that the contact between the two symmetry equivalent molecules is quite close.

4.4.7 Charge

The total charge of A2 was calculated with the program pep_charge (see section 5) (see Fig. 4.23). Under crystallization conditions (pH = 8.2) it has a charge of -2 , which results from the deprotonated acidic functions of the residues Asp2, Glu7 and Cys18 and the protonated N-terminal residue Ala1. As there is also one metal ion (the positively charged sodium atom) and a half acetate molecule in the asymmetric

unit there remains an overall negative charge of 1.5 in the asymmetric unit. But it could be possible that not each Asp2 or Glu7 is deprotonated due to hydrogen bonds to other molecules. However, this could not be detected due to the fuzzy solvent region and the limited resolution of 1 Å of the electron density.

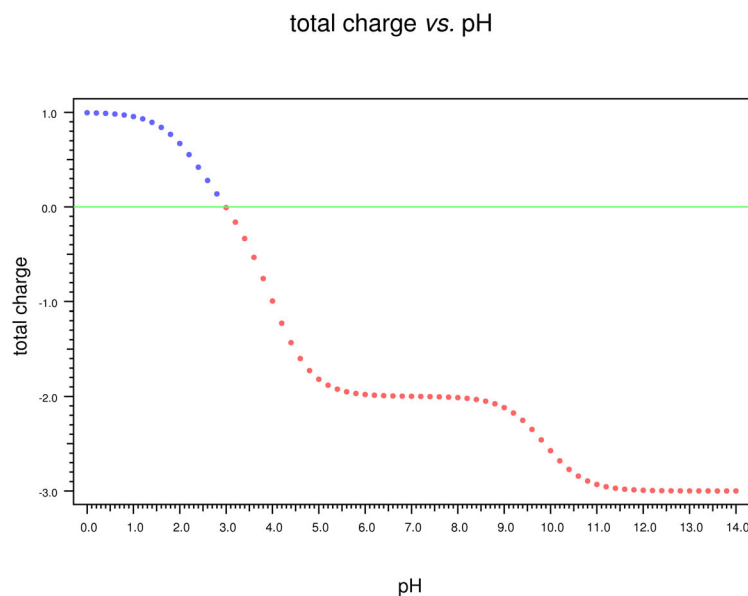


Figure 4.23: Calculated charge of labyrinthopeptin A2 over a pH range of 0.0 to 14.0.

4.4.8 Hydrogen bonds

For investigating the hydrogen network present in A2 a hydrogen bond table was calculated by SHELXL. The calculated angles and distances are shown in Table 4.8.

D-H	d(D-H)	d(H..A)	<DHA	d(D..A)	A
N-H1_1	0.910	2.153	133.50	2.857	O_1009
N-H1_1	0.910	2.478	163.14	3.359	O_1038
N-H2_1	0.910	1.905	152.77	2.746	O_10
N-H3_1	0.910	1.968	156.09	2.824	O_1007
N-H_2	0.880	2.179	133.31	2.856	O_2 $[-x, -y + 1, z]$
N-H_3	0.880	1.875	164.38	2.733	OXT_200 $[-x, -y + 1, z]$
N-H_3	0.880	2.133	165.34	2.993	O_200
N-H_3	0.880	2.569	134.78	3.249	O_200 $[-x, -y + 1, z]$
N-H_3	0.880	2.589	140.43	3.315	OXT_200
NE1-HE1_3	0.880	2.180	156.02	3.005	O_7 $[-x, -y + 1, z]$
NE1-HE1_3	0.880	2.469	113.18	2.929	O_6 $[-x, -y + 1, z]$
N-H_4	0.880	2.208	139.07	2.930	O_2 $[-x, -y + 1, z]$
N-H_5	0.880	2.186	144.11	2.944	O_1004

Table 4.8: Hydrogen bonds in labyrinthopeptin A2.

D-H	d(D-H)	d(H..A)	<DHA	d(D..A)	A
N-H_6	0.880	2.078	167.86	2.944	O_1004
NE1-HE1_6	0.880	2.271	144.66	3.031	O1_18b $[-x + 1/2, y - 1/2, -z]$
N-H_7	0.880	1.943	145.31	2.713	O_1010
N-H_8	0.880	2.161	165.50	3.021	O_1012
N-H_9	0.880	1.898	166.79	2.762	OD1_2 $[-x, -y + 1, z]$
N-H_10	0.880	2.170	149.98	2.965	OD2_2 $[-x, -y + 1, z]$
N-H_11	0.880	2.048	168.85	2.917	OE1_7 $[-x + 1/2, y - 1/2, -z]$
OG1-HG1_11a	0.840	2.481	114.73	2.933	O_1010 $[-x + 1/2, y - 1/2, -z]$
OG1-HG1_11b	0.840	2.225	171.86	3.059	O_8 $[-x + 1/2, y - 1/2, -z]$
N-H_12	0.880	2.066	147.06	2.845	O_1007
N-H_13	0.880	1.906	164.24	2.763	OE2_7 $[-x + 1/2, y - 1/2, -z]$
N-H_14	0.880	2.527	167.66	3.392	O_1001 $[-x + 1/2, y - 1/2, -z]$
N-H_15	0.880	2.951	145.96	3.713	SG_17
N-H_16	0.880	2.080	130.91	2.738	O_1032 $[x, y + 1, z]$
N-H_16	0.880	2.589	177.77	3.468	O_1024
N-H_17	0.880	2.085	167.87	2.951	O_1013
N-H_18a	0.880	2.256	134.23	2.938	O_1013
N-H_18b	0.880	2.204	127.17	2.824	O_1013

Table 4.8: Hydrogen bonds in labyrinthopeptin A2 (cont.).

It can be seen that there are only two intramolecular hydrogen bonds in A2. One bond is between NH_2 and O_10 and the other one is between NH_15 and SG_17. Intermolecular hydrogen bonding occurs much more often between A2 and its symmetry equivalents. Additional to the bonding between and within A2 there is also a hydrogen bonding network to the many water molecules (which are represented by their oxygen atoms and have the residue numbers from 1000 upwards) and the acetate ions (the two acetate oxygen atoms are denoted O_200 and OXT_200, respectively) present in the unit cell.

4.4.9 Crystal packing

A2 crystallizes in the orthorhombic space group $P2_12_12$ with one molecule in the asymmetric unit which makes four molecules in the unit cell. The crystal packing along the a -axis is shown in Fig. 4.24(a) and along the b -axis in Fig. 4.24(b).

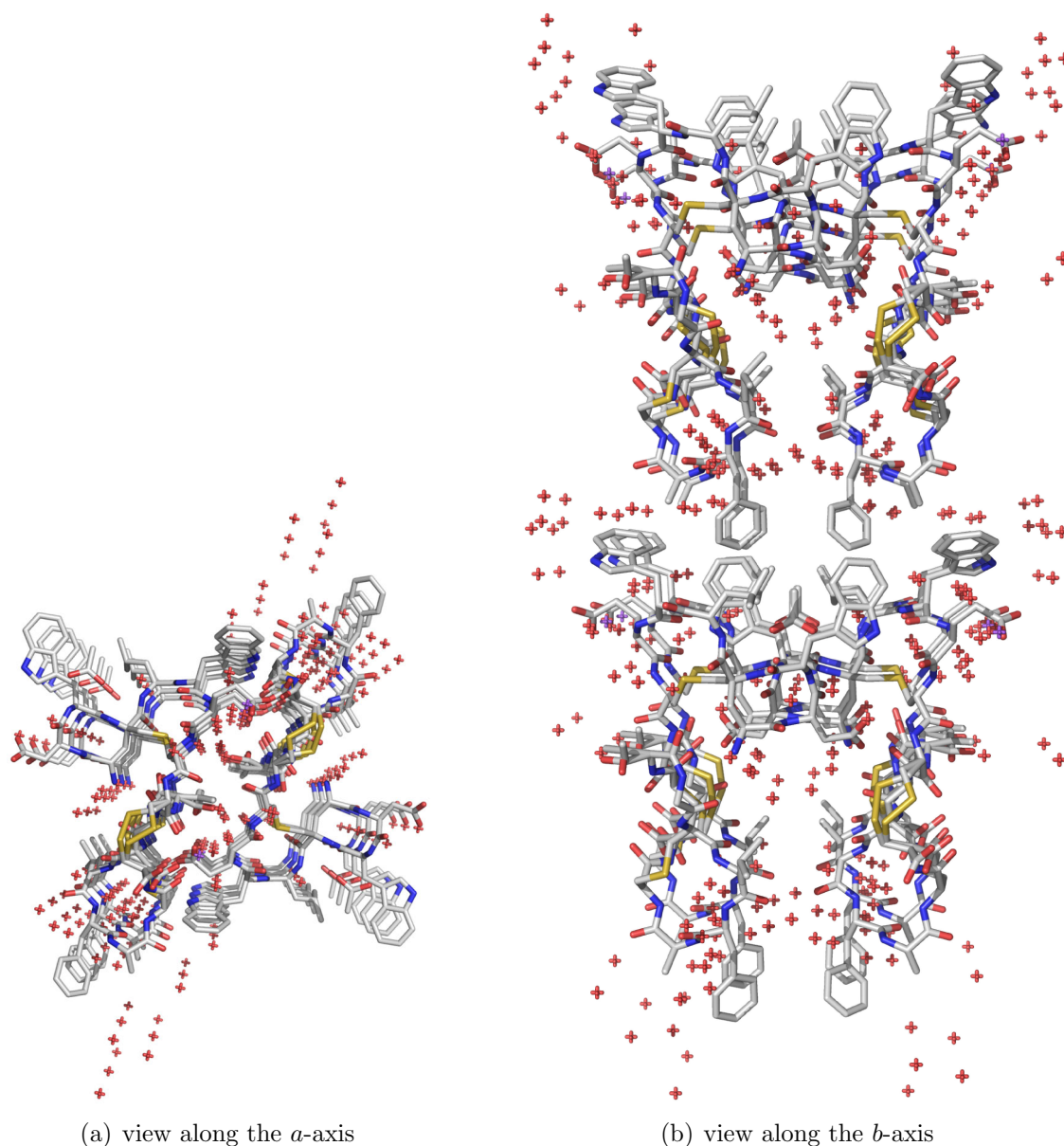


Figure 4.24: Crystal packing in labyrinthopeptin A2.

It can be seen that the molecules are arranged such that channels are formed in which the solvent molecules are distributed. The water molecules arrange quite regularly everywhere in the interspaces in the unit cell but no solvent can be found in the region around the hydrophobic residues Phe14 and Trp3/Leu5.

5 List of developed programs

- **factor_int_sig** reads in a file called `xd_mit_fcalc.hkl` and adds Gaussian noise according to the polar form of the Box-Muller transform to the F_{obs}^2 (which should be ideal data generated by `fco2hkl`); p_1 , μ and σ are variables that can be set to any value; a file called `xd.hkl` is written
- **fco2hkl** reads in a `xd.fco` file and writes a `xd.hkl` format file (`xd_mit_fcalc.hkl`) with F_{calc}^2 as F_{obs}^2 ; if a `xd.hkl` file is present the header is read in and copied to the `xd_mit_fcalc.hkl` file
- **histogram** reads in a `xd_fou.grd` file and counts the frequency of the values; a PostScript file with a histogram representation of the values is printed
- **histomatch_fco** reads in a `xd.fco` file and sorts the F_{calc}^2 as well as the F_{obs}^2 together with their *hkl* indices in descending order; the F_{obs}^2 are replaced by the F_{calc}^2 and written to a `xd.hkl` format file called `xd.hkl_histocorr`; an information on how many F_{obs}^2 have been replaced by their corresponding F_{calc}^2 is printed to the screen
- **invent_hkl** asks the user for the unit cell ($a, b, c, \alpha, \beta, \gamma$), the wavelength λ , lower and upper limits for the resolution (in \AA^{-1}) and the space group; unique reflections for the given resolution range are calculated with intensities and corresponding standard deviations of 1; systematic absences are omitted; the output is written to a `xd.hkl` file called `k2.hkl`
- **jnk2RDA** reads in grid files written by XDFOUR, XDFFT, MoPro or Tonto; it calculates the fractal dimension distribution for the residual density; a more detailed description on the program is given in the text (see section 3.3.1); writes a PostScript file with a graphic representation of the fractal dimension distribution
- **log_histogram** reads in a `xd_fou.grd` file and counts the frequency of the residual density values; calculates the logarithm of the frequency and writes a PostScript file with a graphical representation of the distribution

- **pep_charge** reads in an amino acid sequence in upper case or lower case and the pH, either from a file (the amino acid sequence can span multiple lines, the pH must be given in an extra line) or typed in by hand; calculates the total charge for this sequence and charges for each amino acid; the N-terminus and the C-terminus can have protective groups (entered as '-'); it prints the total charge for the sequence for a pH range from 0.0 to 14.0 in steps of 0.2 to a file called pep_charge.out (space separated); additionally, a PostScript file (called pep_charge.ps) with a graphical output for the total charge *vs.* the pH is written

Bibliography

- [1] W. Massa, "Kristallstrukturbestimmung", Teubner, Stuttgart, **1996**.
- [2] W. H. Bragg and W. L. Bragg, "The Reflection of X-rays by Crystals", *Proc. R. Soc., London* **1913**, *88*, 428-438.
- [3] M. Born, "Quantenmechanik der Stoßvorgänge", *Z. Physik* **1926**, *38*, 803-827.
- [4] D. Sayre, "The squaring method: a new method for phase determination", *Acta Cryst.* **1952**, *5*, 60-65.
- [5] W. Cochran, "A relation between the signs of structure factors", *Acta Cryst.* **1952**, *5*, 65-67.
- [6] W. Cochran, "Relations between the phases of structure factors", *Acta Cryst.* **1955**, *8*, 473-478.
- [7] J. Karle and H. Hauptman, "The phases and magnitudes of the structure factors", *Acta Cryst.* **1950**, *3*, 181-187.
- [8] H. Hauptman and J. Karle, "Structure invariants and seminvariants for noncentrosymmetric space groups", *Acta Cryst.* **1956**, *9*, 45-55.
- [9] J. Karle and H. Hauptman, "A theory of phase determination for the four types of non-centrosymmetric space groups $1P222$, $2P22$, $3P_12$, $3P_22$ ", *Acta Cryst.* **1956**, *9*, 635-651.
- [10] G. M. Sheldrick and T. R. Schneider, "SHELXL: High-resolution refinement", *Methods Enzymol.* **1997**, *277*, 319-343.
- [11] G. M. Sheldrick, "A short history of SHELX", *Acta Cryst.* **2008**, *A64*, 112-122.
- [12] R. F. Stewart, "Valence Structure from Coherent X-Ray Scattering: Fourier Difference Synthesis", *J. Chem. Phys.* **1968**, *48*, 4882-4889.
- [13] R. F. Stewart, "Generalized X-Ray Scattering Factors", *J. Chem. Phys.* **1969**, *51*, 4569-4577.

- [14] R. F. Stewart, "Electron population analysis with generalized x-ray scattering factors: Higher multipoles", *J. Chem. Phys.* **1973**, *58*, 1668-1676.
- [15] R. F. Stewart, "Electron population analysis with rigid pseudoatoms", *Acta Cryst.* **1976**, *A32*, 565-574.
- [16] N. K. Hansen and P. Coppens, "Testing Aspherical Atom Refinements on Small-Molecule Data Sets", *Acta Cryst.* **1978**, *A34*, 909-921.
- [17] A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. R. Mallinson, T. Richter and T. Koritsanszky, "XD2006 - A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental and Theoretical Structure Factors", **2006**.
- [18] R. F. Stewart, "One-electron density functions and many-centered finite multipole expansions", *Isr. J. Chem.* **1977**, *16*, 124-131.
- [19] E. Clementi and D. L. Raimondi, "Atomic Screening Constants from SCF Functions", *J. Chem. Phys.* **1963**, *38*, 2686-2689.
- [20] P. Coppens, "X-Ray Charge Densities and Chemical Bonding", Oxford University Press, New York, **1997**.
- [21] C. B. Hübschle and P. Luger, "Molliso - a program for colour-mapped iso-surfaces", *J. Appl. Cryst.* **2006**, *39*, 901-904.
- [22] D. Leusser, J. Henn, N. Kocher, B. Engels and D. Stalke, "S=N versus S⁺-N⁻: an experimental and theoretical charge density study", *J. Am. Chem. Soc.* **2004**, *126*, 1781-1793.
- [23] S. French and K. Wilson, "On the treatment of negative intensity observations", *Acta Cryst.* **1978**, *A34*, 517-525.
- [24] B. Mandelbrot, "Les objets fractals", Flammarion, Paris, **1975**.
- [25] B. B. Mandelbrot, "The Fractal Geometry of Nature", W. H. Freeman and Company, New York, **1982**.
- [26] K. Falconer, "Fractal Geometry - Mathematical Foundations and Applications", John Wiley & Sons Ltd, Chichester, **2003**.
- [27] G. Cantor, "Ueber unendliche, lineare Punktmannichfaltigkeiten", *Math. Ann.* **1879**, *15*, 1-7.

- [28] G. Cantor, “Ueber unendliche, lineare Punktmannichfaltigkeiten”, *Math. Ann.* **1880**, *17*, 355-358.
- [29] G. Cantor, “Ueber unendliche, lineare Punktmannichfaltigkeiten”, *Math. Ann.* **1882**, *20*, 113-121.
- [30] G. Cantor, “Ueber unendliche, lineare Punktmannichfaltigkeiten”, *Math. Ann.* **1883**, *21*, 51-58.
- [31] G. Cantor, “Ueber unendliche, lineare Punktmannichfaltigkeiten”, *Math. Ann.* **1883**, *21*, 545-591.
- [32] W. Sierpiński, “Sur une courbe dont tout point est un point de ramification”, *C.R. Acad. Sci. Paris* **1915**, *160*, 302.
- [33] W. Sierpiński, “Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée”, *C.R. Acad. Sci. Paris* **1916**, *162*, 629-632.
- [34] H. von Koch, “Sur une courbe continue sans tangente, obtenue par une construction géométrique élémentaire”, *Ark. Mat.* **1904**, *1*, 681-704.
- [35] H. von Koch, “Une méthode géométrique élémentaire pour l’étude de certaines questions de la théorie des courbes planes”, *Acta Math.* **1906**, *30*, 145-174.
- [36] G. Peano, “Sur une courbe, qui remplit toute une aire plane”, *Math. Ann.* **1890**, *36*, 157-160.
- [37] E. W. Weisstein, “Peano curve.”, From *MathWorld* - A Wolfram Web Resource. <http://mathworld.wolfram.com//PeanoCurve.html>.
- [38] http://en.wikipedia.org/wiki/Mandelbrot_set, 08/08/08.
- [39] G. M. Julia, “Mémoire sur l’itération des fonctions rationnelles”, *J. Math. Pure Appl.* **1918**, *8*, 47-245.
- [40] L. F. Richardson, “The problem of contiguity: an appendix of statistics of deadly quarrels”, *General Systems Yearbook* **1961**, *6*, 139-187.
- [41] J. C. R. Hunt, “Lewis Fry Richardson and his contributions to mathematics, meteorology, and models of conflict”, *Annu. Rev. Fluid Mech.* **1998**, *30*, xiii-xxxvi.
- [42] H.-O. Peitgen, H. Jürgens and D. Saupe, “Chaos and Fractals - New Frontiers of Science”, Springer-Verlag, Dordrecht, **2004**.

- [43] B. Mandelbrot, "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension", *Science* **1967**, *156*, 636-638.
- [44] I. N. Bronstein, K. A. Semendjajew, G. Musiol and H. Mühlig, "Taschenbuch der Mathematik", Harri Deutsch, Frankfurt am Main, **1993**.
- [45] F. Hausdorff, "Dimension und äußeres Maß", *Math. Ann.* **1918**, *79*, 157-179.
- [46] B. Guillot, L. Viry, R. Guillot, C. Lecomte and C. Jelsch, "Refinement of proteins at subatomic resolution with MOPRO", *J. Appl. Cryst.* **2001**, *34*, 214-223.
- [47] C. Jelsch, B. Guillot, A. Lagoutte and C. Lecomte, "Advances in protein and small-molecule charge-density refinement methods using MoPro", *J. Appl. Cryst.* **2005**, *38*, 38-54.
- [48] D. Jayatilaka and D. J. Grimwood, "Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography", *Comput. Sci. - ICCS* **2003**, *2660*, 142-151.
- [49] J. Brewer and L. Di Girolamo, "Limitations of fractal dimension estimation algorithms with implications for cloud studies", *Atmos. Res.* **2006**, *82*, 433-454.
- [50] G. E. P. Box and M. E. Muller, "A note on the generation of random normal deviates", *Ann. Math. Stat.* **1958**, *29*, 610-611.
- [51] E. Prince (Ed.), "International Tables for Crystallography, Volume C", Kluwer Academic Publishers, Dordrecht, **2006**.
- [52] G. A. Jeffrey and D. W. J. Cruickshank, "Molecular Structure Determination By X-Ray Crystal Analysis: Modern Methods And Their Accuracy", *Q. Rev. Chem. Soc.* **1953**, *7*, 335-376.
- [53] F. Allen, "A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction", *Acta Cryst.* **1986**, *B42*, 515-522.
- [54] S. Liang, C.-H. Lee, S. I. Kozhushkov, D. S. Yufit, J. A. K. Howard, K. Meindl, S. Rühl, C. Yamamoto, Y. Okamoto, P. R. Schreiner, B. C. Rinderspacher and A. de Meijere, "Bullvalene Trisepoxide and Its Stereospecific Rearrangement to 2,8,12-Trioxahexacyclo [8.3.0.0^{3,9}0^{4,6}0^{5,13}0^{7,11}]tridecane: Two New C₃-Symmetrical Oligocycles with Propeller Chirality", *Chem. Eur. J.* **2005**, *11*, 2012-2018.

- [55] N. Kocher, C. Selinka, D. Leusser, D. Kost, I. Kalikhman and D. Stalke, "Experimental Charge Density Studies of Cyclotetrasilazane and Metal Complexes Containing the Di- and Tetraanion", *Z. Anorg. Allg. Chem.* **2004**, *630*, 1777-1793.
- [56] N. Kocher, "Experimental Charge Density Studies of Highly Polar Bonds", *PhD Thesis*, Bayerische Julius-Maximilians-Universität Würzburg, **2003**.
- [57] U. Shmueli (Ed.), "International Tables for Crystallography, Volume B", Kluwer Academic Publishers, Dordrecht, **2006**.
- [58] J. Drenth, "Principles of Protein X-Ray Crystallography", Springer-Verlag, New York, **1999**.
- [59] K. Y. J. Zhang and P. Main, "Histogram Matching as a New Density Modification Technique for Phase Refinement and Extension of Protein Molecules", *Acta Cryst.* **1990**, *A46*, 41-46.
- [60] V. Y. Lunin, "Electron-Density Histograms and the Phase Problem", *Acta Cryst.* **1993**, *D49*, 90-99.
- [61] H. G. Schlegel, "Allgemeine Mikrobiologie", Georg Thieme Verlag, Stuttgart, **1985**.
- [62] M. T. Madigan, J. M. Martinko and J. Parker, "Brock. Biology of Microorganisms", Prentice Hall, New Jersey, **2000**.
- [63] F. von Nussbaum, M. Brands, B. Hinzen, S. Weigand and D. Häbich, "Antibakterielle Naturstoffe in der medizinischen Chemie - Exodus oder Renaissance?", *Angew. Chem.* **2006**, *118*, 5194-5254.
- [64] G. Jung, "Lantibiotics - ribosomal synthesized polypeptides with sulfide bridges and α, β -didehydroamino acids", *Angew. Chem.* **1991**, *103*, 1067-1084.
- [65] H.-G. Sahl, R. W. Jack and G. Bierbaum, "Biosynthesis and biological activities of lantibiotics with unique post-translational modifications", *Eur. J. Biochem.* **1995**, *230*, 827-853.
- [66] C. van Kraaij, W. M. de Vos, R. J. Siezen and O. P. Kuipers, "Lantibiotics: biosynthesis, mode of action and applications", *Nat. Prod. Rep.* **1999**, *16*, 575-587.
- [67] O. McAuliffe, R. P. Ross and C. Hill, "Lantibiotics: structure, biosynthesis and mode of action", *FEMS Microbiol. Rev.* **2001**, *25*, 285-308.

- [68] G. C. Patton and W. A. van der Donk, "New developments in lantibiotic biosynthesis and mode of action", *Curr. Opin. Microbiol.* **2005**, *8*, 543-551.
- [69] C. Chatterjee, M. Paul, L. Xie and W. A. van der Donk, "Biosynthesis and Mode of Action of Lantibiotics", *Chem. Rev.* **2005**, *105*, 633-683.
- [70] M. J. Horn, D. B. Jones and S. J. Ringel, "Isolation of a new sulfur-containing amino acid (lanthionine) from sodium carbonate-treated wool", *J. Biol. Chem.* **1941**, *138*, 141-149.
- [71] J. M. Willey and W. A. van der Donk, "Lantibiotics: Peptides of Diverse Structure and Function", *Annu. Rev. Microbiol.* **2007**, *61*, 477-501.
- [72] J. Wink, R. M. Kroppenstedt, G. Seibert and E. Stackebrandt, "*Actinomadura namibiensis* sp. nov.", *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 721-724.
- [73] G. Seibert, L. Vértesy, J. Wink, I. Winkler, R. Süssmuth, G. Sheldrick, K. Meindl, M. Brönstrup, H. Hoffmann, H. Guehring and L. Toti, "Antibacterial and antiviral peptides from *Actinomadura namibiensis*", WO 2008/040469 A1, **2008**.
- [74] K. Meindl[#], T. Schmiederer[#], A. Reicke, S. Keller, G. Nicholson, L. Vértesy, J. Wink, M. Brönstrup, G. M. Sheldrick and R. D. Süssmuth, "Labyrinthopeptin A2 - A new lantibiotic with antipain effects in mice", *in preparation*.
[#]These authors contributed equally to the work.
- [75] M. Ries-Kautt and A. Ducruix, *Phase diagrams*, In: Crystallization of Nucleic Acids and Proteins - A Practical Approach, A. Ducruix and R. Giegé (Eds.), Oxford University Press, New York, **1992**, pp. 195-218.
- [76] E. F. Garman and T. R. Schneider, "Macromolecular Cryocrystallography", *J. Appl. Cryst.* **1997**, *30*, 211-237.
- [77] E. Garman, "Cool data: quantity AND quality", *Acta Cryst.* **1999**, *D55*, 1641-1653.
- [78] E. F. Garman and E. P. Mitchell, "Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions", *J. Appl. Cryst.* **1996**, *29*, 584-587.
- [79] E. Garman, "Cool crystals: macromolecular cryocrystallography and radiation damage", *Curr. Opin. Struct. Biol.* **2003**, *13*, 545-551.

- [80] <http://www.jenabioscience.com>.
- [81] Bruker AXS Inc., "SAINT", version 7.23A, Madison (WI), USA, **2006**.
- [82] Bruker AXS Inc., "SADABS - Area detector scaling and absorption correction", version 2007/5, Madison (WI), USA, **2007**.
- [83] Bruker AXS Inc., "XPREP", version 2006/1, Madison (WI), USA, **2006**.
- [84] T. R. Schneider and G. M. Sheldrick, "Substructure solution with SHELXD", *Acta Cryst.* **2002**, *D58*, 1772-1779.
- [85] Bruker AXS Inc., "XP", version 2007/1, Madison (WI), USA, **2007**.
- [86] R. J. Read, "Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors", *Acta Cryst.* **1986**, *A42*, 140-149.
- [87] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics", *Acta Cryst.* **2004**, *D60*, 2126-2132.
- [88] R. A. Engh and R. Huber, "Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement", *Acta Cryst.* **1991**, *A47*, 392-400.
- [89] H. Flack, "On enantiomorph-polarity estimation", *Acta Cryst.* **1983**, *A39*, 876-881.
- [90] IUPAC-IUB Commission on Biochemical Nomenclature, "Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains", *J. Mol. Biol.* **1970**, *52*, 1-17.
- [91] G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations", *J. Mol. Biol.* **1963**, *7*, 95-99.
- [92] R. A. Laskowski, "PROCHECK: a program to check the stereochemical quality of protein structures", *J. Appl. Cryst.* **1993**, *26*, 283-291.
- [93] A. L. Morris, M. W. MacArthur, E. G. Hutchinson and J. M. Thornton, "Stereochemical Quality of Protein Structure Coordinates", *Proteins* **1992**, *12*, 345-364.
- [94] B. Wathen and Z. Jia, "Local and Nonlocal Environments around Cis Peptides", *J. Proteome Res.* **2008**, *7*, 145-153.
- [95] G. N. Ramachandran and A. K. Mitra, "An explanation for the rare occurrence of *cis* peptide units in proteins and polypeptides", *J. Mol. Biol.* **1976**, *107*, 85-92.

-
- [96] D. Pal and P. Chakrabarti, “*Cis* Peptide Bonds in Proteins: Residues Involved, their Conformations, Interactions and Locations”, *J. Mol. Biol.* **1999**, *294*, 271-288.
- [97] M. S. Weiss, A. Jabs and R. Hilgenfeld, “Peptide bonds revisited”, *Nat. Struct. Biol.* **1998**, *5*, 676.
- [98] A. Jabs, M. S. Weiss and R. Hilgenfeld, “Non-proline *Cis* Peptide Bonds in Proteins”, *J. Mol. Biol.* **1999**, *286*, 291-304.
- [99] F. Allen, “The Cambridge Structural Database: a quarter of a million crystal structures and rising”, *Acta Cryst.* **2002**, *B58*, 380-388.
- [100] G. Bunkóczi, M. Schiell, L. Vértesy and G. M. Sheldrick, “Crystal Structures of Cephaibols”, *J. Peptide Sci.* **2003**, *9*, 745-752.
- [101] W. L. DeLano, “PyMOL”, version 0.99rc6, *DeLano Scientific LLC*, Palo Alto (CA), USA, **2006**.
- [102] B. Lee and F. M. Richards, “The Interpretation of Protein Structures: Estimation of Static Accessibility”, *J. Mol. Biol.* **1971**, *55*, 379-400.
- [103] A. Shrake and J. A. Rupley, “Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin”, *J. Mol. Biol.* **1973**, *79*, 351-371.

Appendix

A.1 The net residual electrons in the unit cell

Let the Fouriertransform of the electron density be denoted

$$\begin{aligned}\rho(xyz) &= \frac{1}{V_{UC}} \sum_{hkl} \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \\ &= \frac{1}{V_{UC}} \left(\mathbf{F}_{000} + \sum_{hkl/\{000\}} \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \right)\end{aligned}\tag{A.1}$$

Integration over the unit cell yields:

$$\begin{aligned}\int_{V_{UC}} \rho(xyz) dV &= \frac{1}{V_{UC}} \int_{V_{UC}} \left(\mathbf{F}_{000} + \sum_{hkl/\{000\}} \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \right) dV \\ &= \frac{1}{V_{UC}} (F_{000} V_{UC}) \\ &= F_{000}\end{aligned}\tag{A.2}$$

This equation states that the total scattering power in the unit cell is given by $F(000)$. Accordingly for the residual density:

$$\begin{aligned}\rho_0(xyz) &= \frac{1}{V_{UC}} \sum_{hkl} \Delta \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \\ &= \frac{1}{V_{UC}} \left(\Delta \mathbf{F}_{000} + \sum_{hkl/\{000\}} \Delta \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \right)\end{aligned}\tag{A.3}$$

Integration over the unit cell yields

$$\begin{aligned}
\int_{V_{UC}} \rho_0(xyz) dV &= e_{net}(V_{UC}) \\
&= \int_{V_{UC}} \frac{1}{V_{UC}} \left(\Delta \mathbf{F}_{000} + \sum_{hkl/\{000\}} \Delta \mathbf{F}_{hkl} e^{-2\pi i(hx+ky+lz)} \right) dV \\
&= \int_{V_{UC}} \Delta \mathbf{F}_{000} dV \\
&= \int_{V_{UC}} (\mathbf{F}_{000,a} - \mathbf{F}_{000,b}) dV \\
&= \int_{V_{UC}} \rho(xyz)_a dV - \int_{V_{UC}} \rho(xyz)_b dV \\
&\stackrel{\text{Eq. A.2}}{=} N_a - N_b \\
&\equiv 0
\end{aligned} \tag{A.4}$$

This equation states that the integration over the net residual electrons of the whole unit cell must vanish identically.

A.2 Gross residual electrons in the whole unit cell derived from the Fourier coefficients of the residual density

The number of gross residual electrons is remarkably stable with respect to changes in the grid. Therefore, it might be a good assumption that e_{gross} is a property of the set of ΔF alone. This leads to the question how to calculate

$$e_{gross} = \frac{1}{2} \int_V |\rho_0(xyz)| dV \tag{A.5}$$

from the set ΔF . It will be investigated if the integral of the modulus of $\rho_0(xyz)$, which assumes positive and negative values, can be calculated from the given Fourier

coefficients of its Fourier series. By Parseval's theorem, the modulus squared of any (even complex valued) function can be calculated by summation over the squares of its Fourier series:

$$\int_{V_{UC}} |\rho_0(xyz)|^2 dV = \sum_{hkl} |\Delta F|^2 \quad (\text{A.6})$$

where V_{UC} denotes the volume of the unit cell. If the normalized probability density function (*PDF*) of ρ_0 is known (or can be approximated), *i.e.* the way ρ_0 is distributed in terms of its probability rather than in terms of the spatial distribution, the expectation values can be calculated as follows:

$$\langle |\rho_0| \rangle = \int_{-\infty}^{\infty} PDF(\rho_0) |\rho_0| d\rho_0 \quad (\text{A.7})$$

$$\langle |\rho_0|^2 \rangle = \int_{-\infty}^{\infty} PDF(\rho_0) |\rho_0|^2 d\rho_0 \quad (\text{A.8})$$

The result from Eq. A.8 is identified with the result from Eq. A.6 and e_{gross} is calculated from the ratio of Eq. A.7 and the square root of Eq. A.8. To obtain the desired value for e_{gross} from the set of ΔF a (normalized, as the total probability must be one) Gaussian probability density function for the residual density function is assumed:

$$PDF(\rho_0) d\rho_0 = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\rho_0^2}{2\sigma^2}} d\rho_0 \quad (\text{A.9})$$

The expectation values are calculated:

$$\langle |\rho_0| \rangle = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\rho_0^2}{2\sigma^2}} |\rho_0| d\rho_0 = \sqrt{\frac{2}{\pi}} \sigma \quad (\text{A.10})$$

$$\langle |\rho_0|^2 \rangle = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\rho_0^2}{2\sigma^2}} |\rho_0|^2 d\rho_0 = \sigma^2 \quad (\text{A.11})$$

from which follows

$$\frac{\langle |\rho_0| \rangle}{\sqrt{\langle |\rho_0|^2 \rangle}} = \sqrt{\frac{2}{\pi}} \quad (\text{A.12})$$

For the desired integral it follows that

$$\begin{aligned} \int_{V_{UC}} |\rho_0(xyz)| dV &= \sqrt{\frac{2}{\pi}} \sqrt{\int_{V_{UC}} |\rho_0(xyz)|^2 dV} \\ &= \sqrt{\frac{2}{\pi}} \sqrt{\sum_{hkl} |\Delta F|^2} \end{aligned} \quad (\text{A.13})$$

where the last equation follows from Parseval's theorem. Finally, one obtains:

$$e_{gross} = \frac{1}{\sqrt{2\pi}} \sqrt{\sum_{hkl} |\Delta F|^2} \quad (\text{A.14})$$

Care has to be taken in the evaluation of the sum: the set of ΔF usually comprises the unique reflections whereas the sum must be taken for the Fourier series determining ρ_0 . For example, if the sum runs from $-h_{max}$ to h_{max} , $-k_{max}$ to k_{max} , and from 0 to l_{max} , a factor 2 comes in for the evaluation of the whole sum from the list of unique reflections. In a shorthand notation:

$$\sum_{hkl} \longrightarrow 2 \cdot \sum_{-h_{max}}^{h_{max}} \sum_{-k_{max}}^{k_{max}} \sum_0^{l_{max}} \quad (\text{A.15})$$

Or even shorter:

$$\sum_{hkl} \longrightarrow 2 \cdot \sum_{uq} \quad (\text{A.16})$$

where "uq" abbreviates "unique".

Generalization: The factor $\sqrt{\frac{2}{\pi}}$ connects the integrals over the modulus and the squared modulus of the residual density for the particular case of a Gaussian distribution. As mentioned above, this factor can be calculated for any given distribution, *e.g.* also for Poisson statistics. The main concern here is the Gaussian function for

the reason that a Gaussian function is a good approximation to the limiting case of very good data and a very good model.

Table A.1 shows the respective values for e_{gross} , calculated for a $93 \times 93 \times 107$ grid on simulated data with and without the evaluation of the residual density grid. The results for e_{gross} obtained by explicit calculation of the residual density grid and subsequent summation according to Eq. 3.8 are listed in the second column of the table. The third column shows the result of the summation of the moduli squared of ΔF over the set of unique reflections. The result for e_{gross} from the implicit calculation from the set of $\Delta F = F_{obs} - F_{calc}$ according to the equation given above (Eq. A.14) follows in the fourth column. In this case a factor of 2 in the evaluation of the sum leads to the final equation $e_{gross} = \frac{1}{\sqrt{\pi}} \sqrt{\sum_{uq} |\Delta F|^2}$.

In summary: e_{gross} can be predicted from the set of ΔF without the explicit calculation of a residual density grid. The underlying assumption is a Gaussian residual density distribution.

p_1	e_{gross}^a (e)	$\sum_{uq} \Delta F ^2$	e_{gross}^b (e)	R_1	R_2	$wR_1(w=1)$
0.222	8.2942	218.3111	8.3361	0.0231	0.0180	0.0182
0.333	12.3017	479.1582	12.3499	0.0339	0.0270	0.0270
0.444	16.2007	829.9207	16.2534	0.0443	0.0359	0.0355
0.666	23.6000	1757.4060	23.6516	0.0638	0.0534	0.0517
0.888	30.6022	2953.0915	30.6594	0.0818	0.0705	0.0669
0.999	33.8519	3615.1465	33.9225	0.0899	0.0788	0.0740

Table A.1: Calculation of e_{gross} for the whole unit cell with^a and without^b evaluation of the residual density grid (grid size $93 \times 93 \times 107$) for ideal data on $S(N^tBu)_3$ with noise controlled by p_1 ; the underlying residual density is distributed according to a Gaussian; the index “uq” abbreviates “unique”.

A.3 The gross residual electrons in the unit cell, in a part of the unit cell and standard quality measures

According to the last paragraph, the number of gross residual electrons for the whole unit cell can be calculated from the modulus squared of the Fourier series coefficients of ρ_0 . This suggests a connection between e_{gross} and standard quality measures. In contrast to the quality measures R_1 , R_2 and wR_2 the quality measure wR_1 employs directly the (weighted) sum of residuals, which is minimized during the least squares

refinement:

$$wR_1 = \sqrt{\frac{\sum_{hkl} w |\Delta F|^2}{\sum_{hkl} w F_{obs}^2}} \quad (\text{A.17})$$

Omitting the weights (*i.e.* w is set to 1) and regarding the denominator as a normalization constant C_N , which is fixed for a given set of observed data, this yields

$$R_{LS} = \sqrt{\frac{\sum_{hkl} |\Delta F|^2}{C_N}} \quad (\text{A.18})$$

where the index ‘‘LS’’ abbreviates ‘‘least squares’’. Written in this form the similarity to Eq. A.14 is obvious. It follows that the quality measure $e_{gross}|_V$ (the notation should emphasize for the present paragraph the dependence of the number of e_{gross} on the volume it is evaluated for; therefore, $e_{gross}|_V$ reads: the gross residual electrons in the volume V) is a *local extension* to the quality measure R_{LS} in the following sense: For a given and unaltered set of F_{obs} , the measures $e_{gross}|_{V_{UC}}$ and R_{LS} are linearly dependent. In the same way the number of gross residual electrons is cumulative for N disjoint volumes $V = \sum_i V_i$: $e_{gross}|_V = \sum_{i=1}^N e_{gross}|_{V_i}$ so is the global descriptor R_{LS} composed of local contributions from disjoint parts of the unit cell: $R_{LS}|_V = \sum_{i=1}^N R_{LS}|_{V_i}$

As a consequence, the model for which e_{gross} is at minimum is the same model for which the unweighted square root of the residual sum $\sqrt{\sum_{hkl} |\Delta F|^2}$ is at minimum and e_{gross} is able to decompose the total residual sum into its spatial components.

In summary: $e_{gross}|_V$ is a local and a global measure unlike R_{LS} , which is a global measure, exclusively. For evaluation with respect to the whole unit cell $e_{gross}|_{V_{UC}}$ and R_{LS} describe the same entity in real- and in reciprocal space. For evaluation with respect to a part of the unit cell, $e_{gross}|_V$ effectively describes the additive contribution of that part of the unit cell to R_{LS} (apart from a factor of proportion).

A.4 Prediction of $d^f(0)$ from experimental resolution and grid resolution

By its oscillating nature, a Fourier series terminating with a coefficient c_{max} can generate only a limited number of zero (or any other constant) values. This number is related to the termination order by a factor 2, as a 2π periodic function like a $\sin(x)$ or a $\cos(x)$ wave introduces at maximum 2 zeros, a $\sin(2x)$ or a $\cos(2x)$

introduces at maximum 4 zeros, and a $\sin(nx)$ or a $\cos(nx)$ introduces at maximum $2n$ zeros.

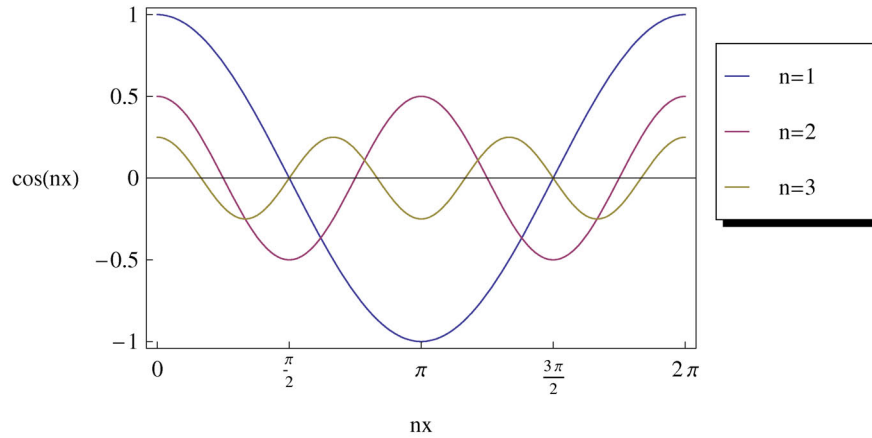


Figure A.1: 2π periodic cosine functions $\cos(nx)$, generating $2n$ zero values; for clarity the functions are scaled by 1, 0.5 and 0.25 for $n = 1$, $n = 2$ and $n = 3$, respectively.

For applications to the residual density analysis this obviously means that the maximum area of *e.g.* the residual density value zero is determined by the experimental resolution. The value zero is chosen as this value occurs most frequently for an ordinary refinement and for the reason that this maximum value does not depend on the variance of the noise. As seen earlier, this value depends (for a perfect model) only on the spatial grid as given by the numbers n_x , n_y and n_z , and by the experimental resolution as given by h_{max} , k_{max} and l_{max} . Furthermore, during a refinement it would be helpful to predict a reasonable value for $d^f(0)$, which can or should be reached under ideal circumstances. Every deviation from this idealized value, may it be positive or negative, can then be interpreted as an introduction of features. The equation is:

$$d^f(0; \varepsilon) = \frac{\log [2h_{max}(n_y - 1)(n_z - 1) + 2k_{max}(n_x - 1)(n_z - 1) + 2l_{max}(n_x - 1)(n_y - 1)]}{\log \left[\sqrt[3]{3(n_x - 1)(n_y - 1)(n_z - 1)} \right]} \quad (\text{A.19})$$

where ε is related to the number of grid points n_x , n_y and n_z (see Eq. 3.12).

A.4.1 Derivation

For the derivation of equation Eq. A.19 it is useful to regard the unit cell as being deformed into a cube and the edges rescaled to unit length. This assumption is not necessary and it does not limit the results in any way, it just simplifies the derivation. The positive integer number h_{max} gives the largest index for h , which, according to

the preceding paragraph, determines the maximum number of zero areas introduced in the direction of the normal vector which points in the x -direction to be $2h_{max}$. As in this case a normalized cube is assumed, this area is $2h_{max}$ times the unit area. This area is covered by $(n_y - 1)(n_z - 1)$ lines which all contribute. The other two terms of the sum in the numerator work analogously for the y - and the z -directions.

In the denominator of Eq. A.19 the number in brackets of which the cubic root is taken just counts the total number of lines existing. The finer the spatial grid is chosen, the larger this number becomes.

A.4.2 Discussion

From Eq. A.19 it can be seen that $d^f(0) = 3$ can be approached, if the spatial grid is chosen in the following way:

$$\begin{aligned} n_x &= 2h_{max} + 1 \\ n_y &= 2k_{max} + 1 \\ n_z &= 2l_{max} + 1 \end{aligned} \tag{A.20}$$

Putting these numbers in Eq. A.19 leads to:

$$d^f(0) = \frac{\log [3(n_x - 1)(n_y - 1)(n_z - 1)]}{\log \left[\sqrt[3]{3(n_x - 1)(n_y - 1)(n_z - 1)} \right]} = 3 \tag{A.21}$$

i.e. for a spatial grid chosen coarse enough a value of $d^f(0) = 3$ can be realized. The coarseness is given by the Nyquist frequency. It is not meaningful to choose a coarser grid because then the spatial resolution is not sufficient to describe all information contained in the experiment. The above equation, however, does not take this case into account. For meaningful results it requires the spatial grid to be at least of the indicated size. *Therefore, if the spatial resolution is chosen according to the Nyquist sampling theorem, $d^f(0) = 3$ holds for the best possible model.* All derivations from this value are due to model errors and data quality limitations or limitations of the applied theory or numerical errors. These introduce the deviations from 3, which manifest themselves as features in the residual density.

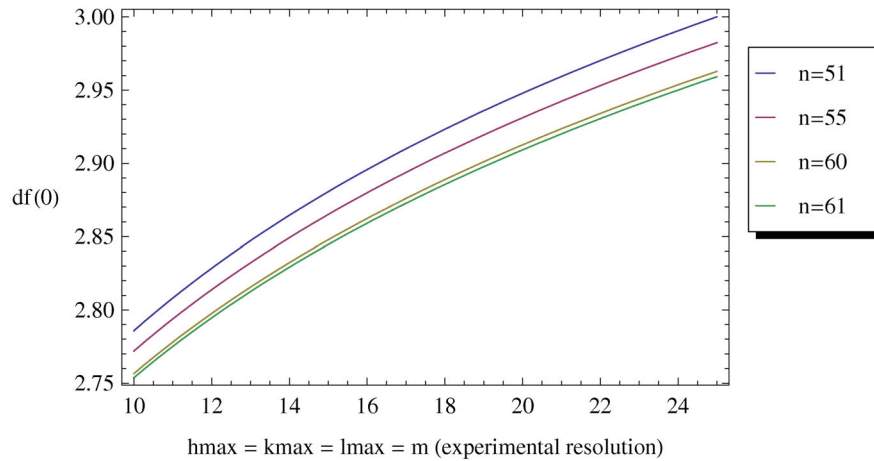


Figure A.2: $d^f(0)$ as predicted by Eq. A.19 for a perfect model and unspecified noise as a function of the experimental resolution $h_{max} = k_{max} = l_{max} = m$ varying between 10 and 25 and for spatial grid resolutions of $n_x = n_y = n_z = n = 51, 55, 60$ and 61 from top to bottom; for $m = 25$ and $n = 51$ the fractal dimension approaches 3; the shape of the curve is concave.

The other asymptotic behavior lies in the limits $\lim_{n_x \rightarrow \infty}$, $\lim_{n_y \rightarrow \infty}$ and $\lim_{n_z \rightarrow \infty}$. For simplicity, it is assumed that $n_x = n_y = n_z = n$ and $h_{max} = k_{max} = l_{max} = m$. Again, the validity of the results is not limited by this assumptions.

$$\lim_{n \rightarrow \infty} d^f(0; n) = \lim_{n \rightarrow \infty} \frac{\log [6m(n-1)^2]}{\log \left[\sqrt[3]{3(n-1)^3} \right]} = 2 \quad (\text{A.22})$$

The value 2 is obtained if the rule of l'Hospital was used.

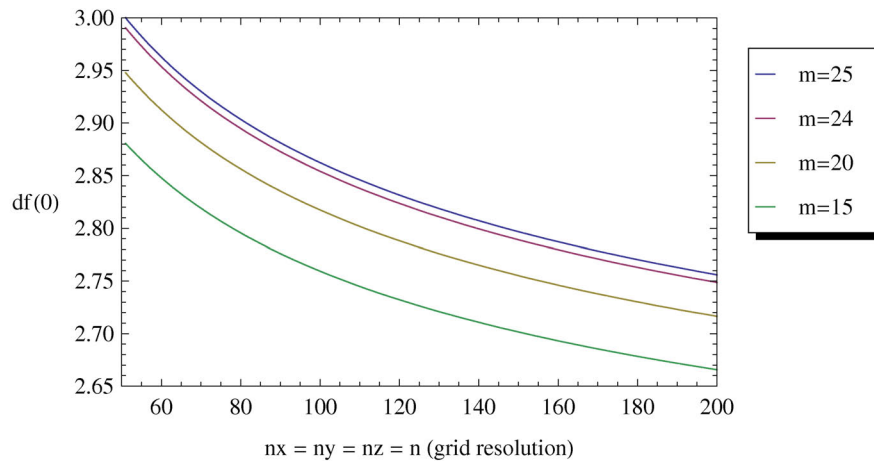


Figure A.3: $d^f(0)$ as predicted by Eq. A.19 for a perfect model and unspecified noise as a function of the spatial grid resolution with $n_x = n_y = n_z = n$ varying between 51 and 200 and for experimental resolutions $h_{max} = k_{max} = l_{max} = m = 15, 20, 24$ and 25 from bottom to top; for $\lim_{n \rightarrow \infty}$ all curves approach $d^f(0) = 2$; the shape of the curve is convex.

Fig. A.4 shows both of the discussed limits in one graph with a convex shape in the direction of increasing spatial resolution and a concave shape in the direction of increasing experimental resolution.

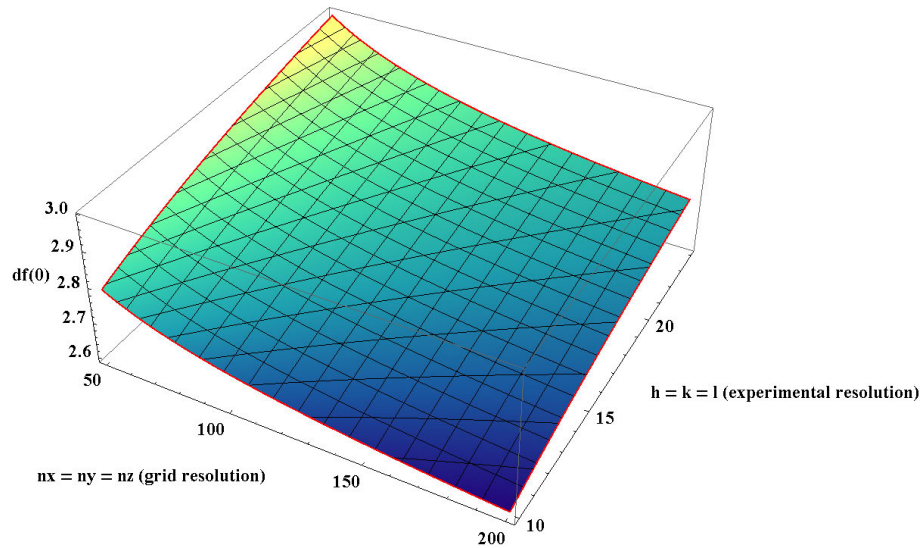


Figure A.4: $df(0)$ as predicted by Eq. A.19 for a perfect model and unspecified noise as a function of the spatial grid resolution with $n_x = n_y = n_z$ varying between 50 and 200 and of the experimental resolution $h_{max} = k_{max} = l_{max}$ varying between 10 and 25; following a line of constant experimental resolution results in a convex line, whereas following a line of constant spatial resolution results in a concave line; the inclined lines indicate contours of constant fractal dimension.

List of Publications

- [1] T. Kurahashi, Y.-T. Wu, K. Meindl, S. Rühl, A. de Meijere, "Cyclopentenones from a Novel [4+1]Cocyclization of Methylenecyclopropanes with Fischer Carbenechromium Complexes", *Synlett* **2005**, 805-808.
- [2] S. Liang, C.-H. Lee, S. I. Kozhushkov, D. S. Yufit, J. A. K. Howard, K. Meindl, S. Rühl, C. Yamamoto, Y. Okamoto, P. R. Schreiner, B. C. Rinderspacher, A. de Meijere, "Bullvalene Trisepoxide and Its Stereospecific Rearrangement to 2,8,12-Trioxahexacyclo [8.3.0.0^{3,9}0^{4,6}0^{5,13}0^{7,11}]tridecane: Two New C₃-Symmetrical Oligocycles with Propeller Chirality", *Chem. Eur. J.* **2005**, *11*, 2012-2018.
- [3] K. Meindl, T. Clark, "Conformations and Tautomers of 5a,6-Anhydrotetracycline", *J. Phys. Chem. B* **2005**, *109*, 4279-4284.
- [4] K. Meindl, L. A. Maslovskaya, M. Noltemeyer, A. I. Savchenko, "3,5,9,12-Tetra-tert-butyl-10,14-dicyclohexyl-2,8-dioxatetracyclo[5.4.0.2^{1,9}.1^{3,7}]tetradeca-4,12-diene-6,11-dione", *Acta Cryst.* **2005**, *E61*, o1950-o1952.
- [5] B. Stulgies, P. Prinz, J. Magull, K. Rauch, K. Meindl, S. Rühl and A. de Meijere, "Six- and Eightfold Palladium-Catalyzed Cross-Coupling Reactions of Hexa- and Octabromoarenes", *Chem. Eur. J.* **2005**, *11*, 308-320.
- [6] S. B. Tsogoeva, M. J. Hateley, D. A. Yalalov, K. Meindl, C. Weckbecker, K. Huthmacher, "Thiourea-based non-nucleoside inhibitors of HIV reverse transcriptase as bifunctional organocatalysts in the asymmetric Strecker synthesis", *Bioorgan. Med. Chem.* **2005**, *13*, 5680-5685.
- [7] S. I. Kozhushkov, T. Preuss, D. S. Yufit, J. A. K. Howard, K. Meindl, S. Rühl, C. Yamamoto, Y. Okamoto, P. R. Schreiner, B. C. Rinderspacher, A. de Meijere, "4,7,11-Triheterotrishomocubanes - Propeller-Shaped Highly Symmetrical Chiral Molecules Derived from Barrelene", *Eur. J. Org. Chem.* **2006**, 2590-2600.

- [8] S. Lösger, O. Schlörke, K. Meindl, R. Herbst-Irmer, A. Zeeck, "Structure and Biosynthesis of Chaetocyclinones, New Polyketides Produced by an Endosymbiotic Fungus", *Eur. J. Org. Chem.* **2007**, 2191-2196.
- [9] T. Kurahashi, S. I. Kozhushkov, H. Schill, K. Meindl, S. Rühl, A. de Meijere, "1,1'-Verknüpfung von Cyclopropan-Derivaten: die helikale Konformation von Quinquecyclopropanol", *Angew. Chem.* **2007**, *119*, 6665-6668; "1,1'-Linked Cyclopropane Derivatives: The Helical Conformation of Quinquecyclopropanol", *Angew. Chem. Int. Ed.* **2007**, *119*, 6545-6548.
- [10] T. Widjaja, L. Fitjer, K. Meindl, R. Herbst-Irmer, "Helical primary structures of 1,2-spiroannulated five-membered rings: attempted synthesis of (\pm)-tetraspiro[4.0.0.0.4.3.3.3]heneicosane", *Tetrahedron* **2008**, 4304-4312.
- [11] K. Meindl, J. Henn, "Foundations of residual-density analysis", *Acta Cryst.* **2008**, *A64*, 404-418.
- [12] G. Seibert, L. Vértesy, J. Wink, I. Winkler, R. Süßmuth, G. Sheldrick, K. Meindl, M. Broenstrup, H. Hoffmann, H. Guehring, L. Toti, "Antibacterial and antiviral peptides from *Actinomadura namibiensis*", WO 2008/040469 A1, **2008**.
- [13] F. Mert-Balci, J. Conrad, K. Meindl, T. Schulz, D. Stalke, U. Beifuss, "Microwave-Assisted Three-Component Reaction for the Synthesis of Pyrido[2',1':2,3]imidazo[4,5-*c*]isoquinolin-5(6*H*)-ones", *Synthesis* **2008**, 3649-3656.
- [14] B. Dittrich, M. A. Spackman, K. Meindl, J. Henn, F. P. A. Fabbiani, "Azulene revisited: Lessons from the classical example of disorder", *in preparation*.
- [15] A. Stasch, S. P. Sarish, B. Nekouieshahraki, H. W. Roesky, K. Meindl, F. Dall'Antonia, T. Schulz, D. Stalke, "Synthesis and Characterization of Alkynyl Complexes of Group 1 and 2 Metals", *in preparation*.
- [16] K. Meindl[#], T. Schmiederer[#], A. Reicke, S. Keller, G. Nicholson, L. Vértesy, J. Wink, M. Brönstrup, G. M. Sheldrick, R. D. Süßmuth, "Labyrinthopeptin A2 - A new lantibiotic with antipain effects in mice", *in preparation*.
[#]These authors contributed equally to the work.
- [17] T. Schulz, K. Meindl, D. Leusser, D. Stern, J. Graf, C. Michaelsen, G. M. Sheldrick, D. Stalke, "A comparison of a microfocuss X-ray source and a conventional sealed tube for crystal structure determination", *in preparation*.

-
- [18] S. Lösgen, T. Bruhn, G. Bringmann, K. Meindl, I. Dix, B. Schulz, A. Zeeck, “(+)-Flavipucine, a configurative puzzle”, *in preparation*.
- [19] K. Meindl, D. Stern, F. Mert-Balci, U. Beifuss, “6-Benzyl-3,4-dimethoxy-9-methylpyrido[2',1':2,3]imidazo[4,5-*c*]isoquinolin-5(6*H*)-one”, *in preparation*.
- [20] N. Kocher, D. Leusser, K. Meindl, K. A. Zachariasse, G. M. Sheldrick, T. Koritsanszky, J. Henn, D. Stalke, “Experimental Charge Density Studies of/with Disordered *N*-Phenylpyrrole and *N*-(4-Fluorophenyl)pyrrole”, *in preparation*.

Poster presentations

- [1] T. Kurahashi, Y.-T. Wu, K. Meindl, S. Rühl, A. de Meijere, “Cyclopentenones from a Novel [4+1]Cocyclization of Methylenecyclopropanes with Fischer Carbenechromium Complexes”, OMCOS13 - IUPAC Symposium on Organometallic Compounds directed towards organic syntheses, Geneva, Switzerland, 17th – 21st July **2005**.
- [2] G. Schwab, D. Leusser, K. Meindl, D. Stalke, “9,10-Bis(diphenylthiophosphanyl)anthracene as a Solid-State Fluorescent Chemosensor”, European Charge Density Meeting - IV, Bollmannsrüh Hotel Brandenburg on the Havel, Germany, 26th – 29th January **2006**.
- [3] K. Meindl, N. Kocher, D. Leusser, H. Ott, D. Stalke, K. Strohfeldt, C. Strohmam, “Understanding Structure/Reactivity Patterns of Lithium Organic Compounds”, Sagamore XV Conference on Electron Charge Spin and Momentum Densities, Bosworth Hall Hotel Warwickshire, UK, 13th – 18th August **2006**.
- [4] K. Meindl, J. Henn, “Fractal dimension as a tool for judging the quality of residual density”, Gordon Research Conference - Electron Distribution & Chemical Bonding (Dynamics And Densities), Mount Holyoke College, South Hadley, MA, USA, 1st – 6th July **2007**.
- [5] K. Meindl, J. Henn, “New Tools for Analyzing the Residual Density”, ECM24, Marrakech, Morocco, 22nd – 27th August **2007**.
- [6] K. Meindl, J. Henn, “Applications of Residual Density Analysis”, 5th European Charge Density Meeting in conjunction with DFG 1178 Annual Meeting, Palazzo Gallio, Gravedona, Italy, 6th – 11th June **2008**.

Curriculum vitae

Persönliche Daten

Name: Kathrin Meindl
Geburtsdatum: 05.11.1976
Geburtsort: Regensburg
Familienstand: ledig

Schulbildung

1983 – 1984 Grundschole Irlbach
1984 – 1987 Volksschule Mitterfels-Haselbach
1987 – 1996 Veit-Höser-Gymnasium Bogen
Juni 1996 Abitur

Hochschulstudium

11/1996 – 11/1998 Grundstudium der Chemie an der Friedrich-Alexander-Universität Erlangen-Nürnberg
11/1998 Diplomvorprüfung
11/1998 – 10/2003 Hauptstudium der Chemie an der Friedrich-Alexander-Universität Erlangen-Nürnberg
11/2002 Diplomhauptprüfung
01/2003 – 10/2003 Diplomarbeit an der Friedrich-Alexander-Universität Erlangen-Nürnberg im Arbeitskreis von Prof. Dr. Tim Clark zum Thema "Konformationen und Tautomerie von Tetrazyklinderivaten"
01/2004 – 10/2008 Promotion an der Georg-August-Universität im Arbeitskreis von Prof. Dr. George M. Sheldrick