THE REFINEMENT OF MACROMOLECULAR STRUCTURES: THEORY AND PRACTICE

by

DALE EDWIN TRONRUD

A DISSERTATION

Presented to the Department of Chemistry
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

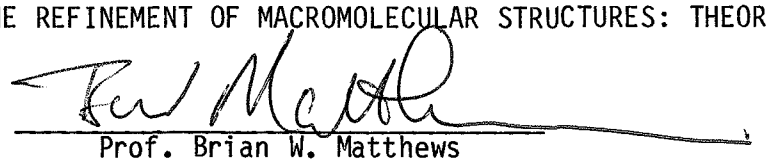
March 1986

An Abstract of the Dissertation of

Dale Edwin Tronrud for the degree of Doctor of Philosophy

in the Department of Chemistry to be taken March 1986

Title: THE REFINEMENT OF MACROMOLECULAR STRUCTURES: THEORY AND PRACTICE

Approved: 

Prof. Brian W. Matthews

The work described in this dissertation encompasses the development of a package of computer programs that facilitate the location and correction of errors in a structural model of a macromolecule. The application of these programs to several different protein structures is also described.

The computer programs were designed to be efficient, simple, and flexible. Their collective function is to optimize the agreement between the experimentally measured X-ray diffraction pattern of a macromolecular crystal and the predicted X-ray diffraction pattern based on a model of that molecule. The programs are sufficiently general that they may be used to refine structures of proteins, nucleic acids, carbohydrates, and lipids.

One application of the program package was to improve the structural model of a bacteriochlorophyll-a containing protein from the green photosynthetic bacterium Prosthecochloris aestuarii. The final model

has a crystallographic R-factor of 18.9% to 1.9Å resolution with good stereochemistry. The procedure of refinement together with examination of the model on a computer graphics system allowed small but significant deviations from planarity in the bacteriochlorophyll-a ring systems to be detected. Further, a prediction of the amino acid sequence, previously unknown, could also be made.

The programs were also used in the determination of the structures of inhibitor complexes of the endopeptidase thermolysin. The inhibitors are peptide analogs in which the planar peptide linkage is replaced by a tetrahedral phosphonamide moiety. This group is a structural analog for the presumed tetrahedral transition-state intermediate produced during the hydrolysis of peptide substrates. Related inhibitors differ from each other by different combinations of chemical substituents near the phosphonamide group. The structures of four complexes were determined and three were refined to at least 1.7Å resolution while the fourth was refined to 2.3Å. From these structures detailed knowledge of the transition state geometry can be inferred and the chemical interactions responsible for binding and catalysis can be identified. It was found that chemically similar inhibitors can, in some instances, adopt very different modes of binding and that these differences in binding geometry are correlated with differences in the kinetics of binding.

VITA

NAME OF AUTHOR: Dale Edwin Tronrud

PLACE OF BIRTH: New Richmond, Wisconsin

DATE OF BIRTH: December 24, 1957

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon
University of Wisconsin-Madison

DEGREES AWARDED:

Doctor of Philosophy, 1986, University of Oregon
Bachelor of Science with honors, 1980, University of Wisconsin-Madison

AREAS OF SPECIAL INTEREST:

Protein Structure Determination Methods
Enzyme Mechanisms
Applications of Computer Graphics

PROFESSIONAL EXPERIENCE:

Research Assistant, Department of Chemistry, University of Oregon, Eugene, 1981-86

Teaching Assistant, Department of Chemistry, University of Oregon, Eugene, 1983-84

Teaching Assistant, Department of Chemistry, University of Oregon, Eugene, 1980-81

PUBLICATIONS:

Holmes, M.A., Tronrud, D.E., & Matthews, B.W. (1983) Structural Analysis of the Inhibition of Thermolysin by an Active-Site-Directed Irreversible Inhibitor. Biochemistry 22, 236-240.

Schmid, M.F., Tronrud, D.E., & Matthews, B.W. (1983) Structural Studies of a Bacteriochlorophyll-Containing Protein. Chemica Scripta 21, 67-69.

Tronrud, D.E., Monzingo, A.F., & Matthews, B.W. (1986) Crystallographic Structural Analysis of Phosphoramidates as Inhibitors and Transition State Analogs of Thermolysin. Eur. J. Biochem., in press.

Tronrud, D.E., Schmid, M.F., & Matthews, B.W. (1986) Structure and X-ray Amino Acid Sequence of a Bacteriochlorophyll a-Protein from Prosthecochloris aestuarii Refined at 1.9A Resolution. J. Mol. Biochem., in press.

ACKNOWLEDGEMENTS

Many people have aided me in the work described in this dissertation. Dr. Lynn Ten Eyck's contributions have been invaluable, both for starting the refinement package project and for leaving the lab before its completion. If he had not done both of these things, I would not have had the opportunity or drive to root through his code and learn about the methods of refinement in trying to interpret it.

While the information flow between Dr. Ten Eyck and myself has been unidirectional, my collaboration with Dr. Hazel Holden has been a lesson in teamwork. The work described in Chapter VI was done with close cooperation between the two of us. I learned the trials and tribulations of data collection with Dr. Holden.

Portions of the work described in Chapter IV were performed by Dr. Mike Schmid, and portions of Chapters V and VI were performed by Dr. Art Monzingo. The work they performed was vital to the completion of those projects.

To develop the refinement package I have had the need to watch people use the programs to see how people solve the problems that arise. This has, in many cases, allowed me to create new mechanisms which make the solution of such problems much easier. I would like to express my gratitude to all those people who have used these programs, both within Brian's lab and those in other labs who have reported back to me. The package would not be nearly as good without them.

I have received funding both from a National Science Foundation grant and from the Molecular Biology Training Grant. Much of the computer equipment was paid for by the Murdock Charitable Trust.

Finally, I must thank Dr. Brian Matthews. He has provided me with a place to work, the facilities with which to work, and the problems on which to work, but more importantly, he has shown me his wisdom and kindness. I cannot imagine a better teacher or boss than Brian Matthews.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

To understand the mechanism of life on the molecular level one must know the structures and fluctuations of the molecules involved. Life is a collection of coupled chemical reactions far from equilibrium. The properties of reactions and the interrelations between them are determined by the chemical nature and orientation of the groups on the surfaces of the molecules involved.

We have not yet reached the point where we can profitably contemplate life in its entirety. Therefore we divide this task into a number of smaller problems. We attempt to understand the mechanism of a particular enzyme, the folding of a protein molecule, or the chemistry of the binding of two macromolecules. From these, more restricted, studies we hope to derive a more general understanding of the entire problem of life.

I have decided to utilize X-ray crystallography in the solution of specific biological problems. X-ray crystallography is a powerful technique that can provide a model for the three-dimensional structure of a macromolecule of interest. Such a model for a protein molecule reveals the overall structure including all the α-helices and β-sheets. In addition, the model also reveals the sites where this molecule can have chemical interactions with other molecules. Once a model for the

structure of a macromolecule has been constructed, crystallographic experiments can often be performed to help understand biological function.

This thesis describes the development of a package of computer programs and, in addition, the application of that package to answer specific questions concerning the structure and function of two different protein molecules. The programs are used to refine an initial crude model of a structure. They continuously vary the position and disorder parameters for each atom in the model in an attempt to maximize the agreement between the observed X-ray diffraction pattern and the diffraction pattern that is predicted from the model. The initial design and some coding was performed by Dr. Lynn Ten Eyck. I completed the coding, made the programs independent of space group, increased the efficiency and wrote the complete documentation. I also added new features, corrected many programming errors, and made the package of sufficient utility that it is now being used in several laboratories around the world. The theory and implementation of these programs are discussed, respectively, in Chapters II and III.

Chapter IV describes the work I performed on the bacteriochlorophyll containing protein. A crude model for this structure had been constructed previously and high resolution diffraction data collected by Dr. M.F. Schmid. A better model was desirable because this protein is one of the few bacteriochlorophyll-containing proteins that is not bound permanently to a membrane. With the exception of the recently-determined structure of a photosynthetic reaction center, the

bacteriochlorophyll protein is the only known structure of a chlorophyll-containing protein. We hoped that an accurate model for the structure of this protein would suggest general rules for the interactions between bacteriochlorophyll groups, including their chlorin rings and their phytol tails, and the interactions between the bacteriochlorophyll molecules and the surrounding protein. This project was complicated by the fact that the amino acid sequence of the protein was not known. This lack of knowledge added the requirement that not only the atoms of the model had to be refined but also the agreement between the model and the data had to be checked often to determine if the sequence of the model was incorrect. This restriction made the refinement more difficult but, on the other hand, when the sequence is known we will also learn much concerning the power and limitations of the crystallographic refinement method. Apart from some preliminary refinement by Dr. Schmid, the high-resolution refinement and analysis of the bacteriochlorophyll protein is my own work.

Chapters V and VI describe the determination and refinement of complexes of inhibitors with the thermostable endopeptidase thermolysin. This protein is a zinc-containing enzyme whose function is to cleave peptide bonds in other proteins. The crystals of the protein diffract X-rays very well and the three-dimensional structure of the protein had been determined previously. The active site is exposed to solvent, allowing even relatively large inhibitor molecules to bind without steric interference from neighboring molecules in the crystal. The inhibitors described here are peptide phosphonamidates and, as such, are

models for the presumed transition state during catalysis. These compounds provide the first experimental evidence showing how the P_2 residue of an extended substrate binds to the thermolysin active site cleft. Some inhibitors bind to thermolysin in the mode predicted for an extended peptide substrate but one inhibitor adopts a different conformation. The structural results help rationalize the observed differences in the kinetics and strength of binding of these inhibitors. The analysis of the thermolysin inhibitors was a joint project involving Drs. A. Monzingo, H. Holden, and myself. I was personally responsible for the high-resolution refinement of most of the thermolysin:inhibitor complexes. In addition, I was extensively involved in the data collection and data processing for three of the complexes.

There is, at the present time, a great deal of interest in the design of drugs. Thermolysin is one of the few systems that allow the methods of drug design to be tested. With this enzyme, one can examine the known structure of the active site, design a compound and predict how strongly it will bind to the enzyme. Through enzyme assays one can then measure the strength of binding and, through crystallography, determine the actual mode of binding. A comparison of the predicted parameters with those observed will give a measure of the reliability of the methods used and clues as to how to improve the method of drug design.

CHAPTER II

DIFFRACTION THEORY

The technique of X-ray diffraction involves directing a beam of X-rays through a crystal of some substance. The resulting pattern of X-rays can be understood as a transformation of the electron density of the crystal. In this chapter I will first describe the relationship between the electron density of the crystal and its diffraction pattern. Then I will discuss the function used to represent the electron density of the crystal and how to derive the diffraction pattern from it. I will conclude with an overview of the methods used to generate a model of the macromolecule's structure.

When X-rays pass through a crystal a diffraction pattern is observed. This pattern is a collection of discrete spots of variable intensity. It has been found that the spots can be represented as an infinite array of evenly spaced points where each point has a value whose magnitude squared is equal to the intensity of the diffraction spot (see Stout & Jensen, 1968). The coordinates used to index this space are called h, k, and l, and together are usually referred to as the vector $\underline{s}$. The space itself is called reciprocal space. Why this space does not contain the intensities directly will be made clear later in this chapter.

Reciprocal space is related to the electron density of the crystal by a mathematical operator called the Fourier transform (T[]). The operation performed by the Fourier transform is to change to the basis set of a function from x, y, and z, to the infinite set of complex exponentials given in (2.1).

$$\exp(2\pi i\ (hx + ky + lz)) \tag{2.1}$$

These functions are nothing more than a set of sine waves of variable frequencies determined by h, k, and l. Therefore the Fourier transform converts a spatial distribution to a frequency distribution. Its most familiar use is the conversion of the complicated waveform of sound into a frequency spectrum.

The Fourier transform operator has a number of interesting and useful properties. First of all it is a linear operator. This means that the operator commutes with addition. The convolution theorum says that the Fourier transform of the product of two functions is the convolution of the Fourier transform of each, and that the Fourier transform of the convolution of two functions is the product of their transforms. Gaussians have particularly simple Fourier transforms. These rules are shown in equations (2.2), (2.3), (2.4), and (2.5).

$$T[f(\underline{r}) + g(\underline{r})] = T[f(\underline{r})] + T[g(\underline{r})] \tag{2.2}$$

$$T[f(\underline{r})g(\underline{r})] = T[f(\underline{r})] * T[g(\underline{r})] \tag{2.3}$$

$$T[f(\underline{r}) * g(\underline{r})] = T[f(\underline{r})]\ T[g(\underline{r})] \tag{2.4}$$

$$T[\exp(-br^2)] = \left(\frac{\pi}{b}\right)^{3/2} \exp(-\pi^2 s^2/b) \tag{2.5}$$

A function that is useful and also has a simple Fourier transform is the Dirac δ function ($\delta(\underline{r})$). This function is defined in equation (2.6) and some of its properties are given in (2.7), (2.8), and (2.9).

$$\delta(\underline{r}) = \lim_{\alpha \to 0} \frac{1}{\alpha} \exp(-\pi r^2/\alpha^2) = \begin{cases} 0 & \underline{r} \neq (0,0,0) \\ \infty & \underline{r} = (0,0,0) \end{cases} \tag{2.6}$$

$$\int_{-\infty}^{\infty} \delta(\underline{r}) d\underline{r} = 1 \tag{2.7}$$

$$T[\delta(\underline{r}-\underline{r}_0)] = \exp(2\pi i(\underline{s}\cdot\underline{r}_0)) \tag{2.8}$$

$$\delta(\underline{r}-\underline{r}_0) * g(\underline{r}) = g(\underline{r}-r_0) \tag{2.9}$$

Because it changes the origin of any function with which it is convuluted (2.9), the Driac δ function is very useful for shifting functions that are defined with one origin but need to be used with another origin.

A lattice is a repeating array of δ functions and will be represented as $\Lambda(\underline{r})$. A particular lattice is described by the distances between δ functions in the three directions of space, and the angles between neighboring δ functions. These parameters are called lattice constants. One fact has yet to be stated: the Fourier transform of a lattice is another lattice whose lattice constants are different from, but related to, the lattice constants of the source function.

The Electron Density Model

A crystal is composed of a collection of molecules called a unit cell that is repeated by translation to fill all of space. Each

molecule is composed of atoms. Normally the electron density of the atoms is represented as a function, with spherical symmetry, which depends on the elemental type of the atom. The parameters used to describe the electron density of an atom are the position of the center of the atom, the "temperature factor", which is a measure of the motion of the atom, and the "fractional occupancy" of the atom. The temperature factor is used to model vibration about the center of the electron cloud, while the occupancy is used to model static disorder. The electron density for an atom is calculated using equation (2.10).

$$\rho_i(\underline{r}) = \delta(\underline{r}-\underline{x}_i) * T^{-1}[O_i \exp(-B_i s^2/4) f_i(\underline{s})]$$
$$\text{where } f_i(\underline{s}) = A \exp(-Bs^2/4) + C \exp(-Ds^2/4) + E \qquad (2.10)$$

O_i is the fractional occupancy for atom i, $\underline{x}_i$ is the position vector of the center of the atom, and B_i is the temperature factor. The second part of the convolution defines the electron density of the atom if it was located at the origin of the coordinate system. This function is convoluted with a δ function to move it to the proper location in the unit cell.

$f_i(\underline{s})$ is called the scattering factor for the atom. It is the Fourier transform of the electron density of a stationary atom of the same elemental type as atom i. To define $f_i(\underline{s})$ a quantum mechanical calculation is performed to determine the wave functions which describe the atom. Then the density calculated from those wave functions is spherically averaged and an electron density profile is produced and Fourier-transformed to obtain the scattering of the atoms as a function

of scattering angle (International Tables for X-ray Crystallography, Vol. IV, 1974). The values can be approximated by the sum of two gaussians and a constant (Forsyth & Wells, 1979). One set of constants has been determined for each element, and usually a different set is calculated for each oxidation state of the various metal ions.

The electron density for the atom when at the origin is defined in terms of its Fourier transform and, therefore, an inverse Fourier transform has entered the Equation (2.10). This is done because the temperature factor parameter (Bi) is particularly easy to define in reciprocal space. The inverse Fourier transform behaves in the same manner as the regular transform and can, in general, be calculated using only forward transforms (Equation 2.11).

$$T^{-1}[f(s)] = T[f(-s)] \qquad (2.11)$$

In Equation (2.10) several approximations are being made. The first is that the effect of the motion of the atom can be modeled by a Gaussian factor. This has been shown to be true for the case of isotropic, harmonic vibration (see Stout & Jensen, 1968). It is not necessarily true for all of the real effects that temperature factors attempt to account for in protein crystal. Equation (2.10) is used because it is accurate for vibrations of small amplitude, and Gaussians have simple Fourier transforms. The form of the scattering factor assumes that all atoms are spherically symmetric and that the shape of the electron density is not affected by the chemical state of the atom.

Equation (2.10) requires five parameters for each atom. One usually does not have enough diffraction data to determine that many

parameters when working with macromolecules. Therefore it is customary to hold the fractional occupancies fixed at full occupancy. If the crystal does not diffract beyond a resolution of 2.5Å there normally are not enough data to determine even the four remaining parameters. In such cases groups of atoms are required to have the same value for their temperature factors and, in the limiting case, all the atoms are forced to have the same temperature factor. If even fewer data are available restrictions would have to be made to the freedom of the positional parameters.

The electron density for the entire unit cell is modelled as the sum of the electron density of all the atoms in the unit cell (2.12).

$$\rho_{uc}(\underline{r}) = \sum_{i}^{atoms} \rho_i(\underline{r}) \qquad (2.12)$$

Usually the macromolecule grows in a crystal form that has some symmetry within the unit cell. These symmetry operators allow the entire contents of the unit cell to be described completely by defining the parameters for only a subset (the "asymmetric unit") of the total number of atoms.

To generate the electron density for the entire crystal from the electron density of the unit cell we must represent mathematically the duplication, translation, and addition of the unit cells. These operations can best be represented as convolution of the unit cell with a lattice. To generate the entire crystal from a single unit cell of density, we require a lattice with a δ function at the origin of each unit cell in the crystal. This lattice is called the "real lattice" and

its lattice constants are called the cell constants. The convolution of the lattice of δ's with the unit cell density is the function for the electron density for the entire crystal (2.13).

$$\rho(\underline{r}) = \Lambda_{\text{real}}(\underline{r}) * \rho_{uc}(\underline{r}) \tag{2.13}$$

The Diffraction Pattern

As stated earlier the diffraction pattern is related to the Fourier transform of the electron density of the crystal. We now have a function which describes this electron density and because we have carefully chosen our function, we can analytically calculate the Fourier transform of this function (2.14).

$$F(\underline{s}) = \Lambda(\underline{s}) \sum_{i}^{\text{atoms}} O_i \exp(-B_i s^2/4) f_i(\underline{s}) \exp(2\pi i \underline{s}\cdot\underline{x}_i) \tag{2.14}$$

Equation (2.14) is the product of a lattice and a function. One of the properties of a lattice is that such a product is zero everywhere except where the δ functions are non-zero. At those points the product gives the value which that point would have if there was not lattice function. In effect, the product is just the second function sampled only at the points of the reciprocal lattice. The units of h, k, and l are chosen so that this situation occurs only for integral values of the coordinates. For those points Equation (2.14) reduces to Equation (2.15) and at all other points reduces to zero.

$$F(\underline{s}) = \sum_{i}^{\text{atoms}} O_i \exp(-B_i s^2/4) f_i(\underline{s}) \exp(2\pi i \underline{s}\cdot\underline{x}_i) \tag{2.15}$$

The discrete points of $F(\underline{s})$, called the "structure factors", are complex numbers. The intensities of the spots in the diffraction pattern are proportional to the square of the magnitude of the corresponding structure factor.

Early Stages of Structure Determination

In the previous sections of this chapter a method of calculating the structure factors from the parameters of the model for the macromolecule has been developed. In the early stages of structure determination there is no model for the macromolecule with which to work. Knowing the relationship between reciprocal space and real space, Equation (2.16) can be written.

$$\rho(\underline{r}) = T^{-1}[\,|F(\underline{s})| \exp(i\,\phi(\underline{s}))] \qquad (2.16)$$

The magnitude of $F(\underline{s})$ is determined from the diffraction data but the phase angle ($\phi(\underline{s})$) cannot be directly measured from the data. If experiments can be performed which will provide an estimate of the phase angle, an electron density "map" can be calculated. Such a map will sample the real space unit cell at discrete locations. Each location has a value representing the electron density at that point. Such a map can be examined by several techniques and a crude model for the structure of the molecule can be constructed.

Several methods are now available which can be used to derive estimates of the phases of the structure factors. These methods include multiple isomorphous replacement (MIR), anomalous dispersion,

non-crystallographic symmetry averaging, and molecular replacement. Most of these methods use information about the structure beyond the diffraction pattern of the normal crystal. The MIR technique involves the measurement of diffraction pattern changes caused by the binding of heavy metal atoms to the macromolecule. Molecular replacement can be used if the macromolecule is believed to be similar to a known structure. If this is the case, phases can be calculated from the known structure and used to calculate an electron density map for the unknown structure.

Little of my work has dealt with projects that have used these methods and I will not describe them in detail. It is sufficient to say that such methods exist and that a crude model for the macromolecule can be constructed. However, a need exists to adjust the parameters of the crude model to achieve the best possible agreement between the diffraction pattern predicted by the model and the pattern actually observed. Much of my dissertation is devoted to this problem.

Macromolecular Structure Refinement

Refinement is the process whereby the parameters of a model are adjusted to improve the fit of the model to the observed data. This usually involves a least-squares minimization of a residual between the observed data and the values predicted from the model.

Crystallographic refinement became a part of the macromolecular structure solution methods about 10 years ago. At that time traditional crystallographic refinement was first applied to a protein structure

(Watenpaugh et al., 1973). It rapidly became clear that, while this method worked well for the case of a small protein which diffracted to high resolution, least square minimization against crystallographic data would not work when applied to the more typical case of a large protein structure whose outer diffraction limit lies between 3Å and 1.5Å. This problem seems to arise simply from a lack of sufficient number of observations.

A typical protein structure will have between 5,000 and 15,000 parameters, three positional and an isotropic temperature factor for each atom. However, there are generally only about 10,000 to 40,000 observed reflections in the data set. The critical parameter is the number of observations per parameter. For proteins this number ranges from 2 to 3, while for "small molecules" the number often is as large as 10 to 20. Therefore the basic problem with protein refinement is a lack of observations.

The most commonly used source of additional information about a protein structure is the knowledge about their stereochemistry. The distances between pairs of atoms, the angles between triplets of atoms, and the closest non-bonded distances allowed between atoms are all well known. The fact that certain groups of atoms are arranged in planes can also be used. Most of the macromolecular refinement programs in use today incorporate this information into the refinement process (Hendrickson & Konnert, 1980; Jack & Levitt, 1978).

However, it has been found profitable to add even more information, if more information is known. Co-refinement of X-ray and Neutron data

for the BPTI structure (Wlodawer et al., 1984) has generated a model whose phases can bring forth new information from difference maps generated from the X-ray data alone. This result implies that X-ray refinement with stereochemical restraints cannot extract all of the information which lies within the X-ray data. Additional data, when properly used, can extract additional information.

The computer program which I developed, in conjunction with Dr. Ten Eyck, was designed to make the incorporation of additional data very easy. It will modify the parameters of a model to improve its agreement with both the observed diffraction pattern and the principles of stereochemistry. In addition, the package was designed so that additional programs, dealing with other sources of information, could be added without any modification of the existing programs. For example, a program has been written which will allow the incorporation of information from experimentally determined phase angles. A full description of the macromolecular refinement package will be the subject of the following chapter.

CHAPTER III

AN EFFICIENT GENERAL-PURPOSE LEAST-SQUARES REFINEMENT PROGRAM FOR MACROMOLECULAR STRUCTURES

Abstract

A package of programs has been developed for efficient, restrained, least squares refinement of macromolecular crystal structures. The package has been designed to be as flexible and general-purpose as possible. The process of refinement is divided into basic units and an independent computer program handles each task. Each functional unit communicates with other programs in the package by way of files of well-defined format. To modify or replace any program requires that the user need only understand the function of that particular element. Stereochemical restraints are defined in a general way that can be applied to proteins, nucleic acids, prosthetic groups, solvent atoms and so on. Guide values for bond lengths and bond angles are specified in a straightforward, direct manner.

In order to make the package as efficient as possible, the fast Fourier transform algorithm is used for all the crystallographic transformations. To highlight potential errors in the refined structure the user can list those atoms that have the worst bond lengths and angles, or have the largest positional, temperature factor or occupancy gradients. It is also possible to check that protein and solvent atoms

do not sterically clash with symmetry-related neighbors. Applications of the program package to a bacteriochlorophyll-containing protein, thermolysin-inhibitor complexes and mutants of bacteriophage T4 lysozymes are described.

Introduction

There are a number of potential difficulties with the refinement of macromolecular structures including the unfavorable ratio of observations to parameters, the magnitude of the computational requirements, and deficiencies in the starting model ranging from small errors in the coordinates to gross errors arising from misinterpretation of the electron density map. These difficulties have led to the development of different refinement strategies, each of which has its own advantages and disadvantages (e.g. see Diamond, 1971; Watenpaugh et al., 1973; Freer et al., 1975; Sussman et al., 1977; Jack & Levitt, 1978; Konnert & Hendrickson, 1980; Agarwal, 1978; Jones & Liljas, 1984). In order to increase the number of observations it is usual to include knowledge of the stereochemistry of the protein. Bond distances, bond angles, planarity and limits on the approach distances of non-bonded atoms can all be specified. It can also be profitable to incorporate additional information, as in the co-refinement of bovine pancreatic trypsin inhibitor with X-ray and neutron data (Wlodawer & Hendrickson, 1982). If other data are available, such as, for example, independent phase information from isomorphous replacement or anomalous scattering, or phase information from molecular replacement, it might be desirable to include this information as well.

The package of programs described uses the principle of restrained least-squares refinement. The package is designed to be as general-purpose as possible. Stereochemistry, for example, is defined in a general way that can be applied to proteins, nucleic acids, prosthetic groups, solvent atoms and so on. Also the package has been made as efficient as possible by using the fast Fourier transform algorithm to carry out all the crystallographic transformations.

One limitation of many refinement programs is their inflexibility; portions cannot be replaced nor new functions added without extensive modification of the existing code. This limits the ability of the user to experiment with different refinement strategies. In order to modify the refinement program one must understand the data structure and algorithms of the entire program. The refinement package described here was designed to avoid this limitation. The process of refinement is broken down into basic units and an independent computer program handles each task. Each functional unit communicates with the other programs in the package by way of files of well defined format. To modify or replace any program only requires that the user understand the function of that program, the rest of the programs will function as before. In this manner calculations which can be optimized by space group specific algorithms (such as fast Fourier transforms) can be calculated differently for different crystal structures by a simple substitution of the appropriate program.

Features and Organization

The package of computer programs was designed to meet five major objectives which will be briefly discussed below.

(i) It should be possible to replace existing functions or add new functions without modifying existing code.

(ii) The programs should have a common mechanism for reading data.

(iii) It should be easy to define standard geometry for new and unusual chemical groups.

(iv) It should be possible to constrain specified groups of atoms to behave as rigid bodies or to be held constant during refinement.

(v) The program should provide tools to aid the user in the detection of errors in the model that are beyond the ability of the refinement package to correct.

Overall Organization

The need to partition a refinement program into independent, functional units was mentioned in the introduction and has shaped the overall organization and structure of the package. The different functions that are minimized in the refinement are treated as separate "terms" where each term is defined on the basis of the calculations required to evaluate the term and its gradient. Most commonly, two terms are included: a crystallographic term and a stereochemical term. The programs required to calculate a term and its gradient are collectively referred to as a "module". The overall refinement package

consists of the control program plus a variable number of modules. The control program combines the information presented by all the modules to determine the direction in which to shift the parameters of the model, and, eventually, to determine the magnitude of the shift. This program "knows" nothing about specifics of the terms that are handled by the various modules. The stereochemical module is implemented as a single program. Because it involves several Fourier transforms, the crystallographic module is broken up into five different programs. This fractionation allows appropriate space-group-specific fast Fourier transform (FFT) programs to be used for different projects.

Data Input and Data Transfer

Because of the number and independence of the programs in the package it is very desirable that the input format for all programs be the same. Although some data files are created by the computer and others by the user, the style of input has been designed with emphasis on the benefit of the user. All input is token based. The input is read in free-format, each token being separated from the next by a delimiter such as a space or a comma (see Fig. 1, Fig. 2, and Table 1 for examples). The first token on the card is the keyword. Keywords fall into two classes: data keywords and command keywords. The order of data cards is unimportant. When a command card is encountered, the required operation is performed on whatever data has been read to that point. Input and output is handled by a set of library routines which perform the basic operations of reading in cards, separating tokens, and building numbers from particular tokens.

```
<Atom card>      :==
    <Atom keyword> <Atom type> <Atom parameters>
                   <Atom name> <Residue name> <Chain name>

<Chain card>     :==
    CHAIN <Chain name> <Chain type>
          {<Chain name>" | "<Residue name> <Linkage type>}

<Residue card>   :==
    RESIDUE <Chain type>" | "<Residue name> <Residue type>
            {<Residue name> <Linkage type>}

<Geometry card> :==
    GEOMETRY <Cluster type> <Restraint type> <Standard value>
             <Value's sigma> <Atom name> {<Atom name>}
```

Where:

```
<Atom keyword>     :== ATOM | ATOMC | ATOMG

<Cluster type>     :== <Residue type> | <Linkage type>

<Restraint type>   :== BOND     | ANGLE | TORSION |
                       TRIGONAL | PLANE | CHIRAL
```

The nomenclature is:

```
<name>    is a word or number
A | B     means A or B
{ A }     means A is repeated 0 or more times
" | "     means that the | is to be taken literally
```

FIGURE 1. The Definition of the General Data Cards

These are the definitions for the cards used to define the standard geometry for a molecular model. Each coordinated card contains the atom's name as well as the names of the residue and chain in which it resides. Each chain must have its type defined on a CHAIN card. The sequence and connectivity of that type of chain must be defined on several RESIDUE cards and the restraints associated with each residue type and linkage type are defined on GEOMETRY cards.

A related objective was to allow the constituent programs to be easy to write and understand. Sets of common routines have been placed in libraries that are used by most of the programs. These libraries contain routines that process data cards and build the internal data structures. Also there are other routines that locate required information within the data structure. By not having to rewrite these routines one can implement a new program in a very short time and, by having a common internal structure, the existing programs can be understood more easily. On the other hand, although these library routines are available, they do not have to be used when one wishes to add a new program to the refinement package.

Geometry Definition

Because one often needs to include unusual inhibitors or cofactors in the refinement, it is very desirable that the definition of geometry should be general. Often, the structures of these small molecules have not been determined and their "ideal" geometry must be constructed from fragments whose structures are known. In the present package, standard geometry is defined by breaking the structural model into components, such as amino acids or nucleotides or cofactors. The geometry restraints are then defined in a general way for each component and for the linkages between components.

There are two ways in which stereochemical information can be incorporated into the refinement process; the information can be added as additional observations (restraints) (cf. Hermans & McQueen, 1974;

Konnert & Hendrickson, 1980) or the model can be parameterized (constrained) in such a way that the stereochemistry is always "ideal" (cf. Diamond, 1966; Warme et al., 1972). Although the use of constraints rather than restraints does result in a more favorable ratio of observations to parameters, we have chosen the latter approach. There are several reasons for this, the first being that a restrained model is physically more realistic than a constrained one, as constraints are usually implemented (e.g. see Ten Eyck et al., 1976). A second reason is that different types of restraint can be individually weighted. This allows the user to put a smaller weight on geometric restraints for which the "ideal" values are uncertain. Another reason is that the significance of deviations from ideality can be evaluated by calculating the standard deviation of all the observations within the same class of restraint. At the same time the library of standard values can be tested for accuracy. If there is an error, then there will be a systematic difference between the library value and the value obtained from the refined model. (The "stereochemistry" program in the refinement package will perform both the above tests.) Finally, the use of restraints allows the stereochemical information to be incorporated into the refinement in a manner formally similar to the incorporation of the crystallographic observations (see below). This allows simpler and more efficient code.

Rigid Body Refinement

It is often desirable to have the option of holding portions of the structure fixed, or constraining a group to move as a rigid body (e.g.

see (Sussmann et al., 1977). Because these options do not depend on the nature of the function being minimized they have been implemented in the control program.

The control program also allows one to limit the range of values that the temperature factors and occupancies can assume and, in addition, to constrain a group of atoms to have the same temperature factor or occupancy (e.g. to allow for partial occupancy by an inhibitor).

Detection of Errors in the Model

Finally, in order to highlight potential errors in the current model of the structure, each module can list those atoms that most seriously violate the restraints of the refinement. The stereochemistry module lists, for each class of geometrical restraint, the worst discrepancies between the model and the "ideal" values. Similarly, the crystallographic module can list those atoms with the largest positional, temperature factor or occupancy gradient. Experience has shown that these lists are particularly helpful in pointing out areas of the model that are likely to be in error and may need to be corrected by manual adjustment of the model.

Theoretical Background

Our goal is to minimize a suitable function of the observations in terms of a structural model specified by variables such as coordinates, thermal factors and occupancies. The function used in least squares

refinement is

$$M = \sum_j W(j)\,[Q_o(j) - Q_c(j,\underline{p})]^2 \qquad (3.1)$$

where $Q_o(j)$ is the experimental value for observation j, $Q_c(j,\underline{p})$ is a corresponding value calculated from the structural model, $\underline{p}$, and W(j) is the desired weighting function. The sum in Equation (3.1) is over all observations, but can be separated into different terms based, for example, on the crystallographic observations $\underline{s}$ and the stereochemical observations b (Equation 3.2) (see Appendix B for additional details).

$$M = \sum_{\underline{s}} W(\underline{s})[Q_o(\underline{s}) - Q_c(\underline{s},\underline{p})]^2 + \sum_b W(b)[Q_o(b) - Q_c(b,\underline{p})]^2 \qquad (3.2)$$

More terms could be added if other classes of observation were available. The gradient of M can also be separated into similar terms. This means that the calculations for the crystallographic term can be kept completely separate from calculations for the other terms.

The computational problem is to determine a set of parameters which minimizes M. There exist function minimization methods which use no derivatives, which use only first-derivatives, and which use second-derivatives, in order of increasing power of convergence and increasing computational cost. In the present case there are several reasons for using first-derivative methods.

(i) The radius of convergence of first-derivative methods is larger than that of second-derivative methods, and in these problems one often starts far from the minimum.

(ii) The computational cost of first-derivative methods is proportional to N (the number of parameters) instead of N^2. For large N this is very important.

(iii) Implementation of parameter constraints for holding variables constant, or for requiring variables to behave as rigid groups, is particularly simple for first-derivative methods. Formulate the new parameters as linear combinations of the old, and edit the list of derivatives.

First-derivative methods all use the same general strategy, namely, calculation of the shift direction followed by a line search for a minimum in the chosen direction. The present package uses the conjugate gradient method (Fletcher & Reeves, 1964). In this procedure the changes in the gradient vector from cycle to cycle are used to approximate the second derivative without actually having to compute this quantity.

By using the method of Agarwal (1978) the amount of computer time required to calculate the gradient of the crystallographic term is only slightly longer than the calculation of an FFT of the structure. The time to calculate the gradient of the stereochemical term is, in comparison, miniscule. The stereochemical and crystallographic gradients are combined with the shift vector of the previous cycle to give the direction (but not the magnitude) of the shift for each parameter. The search along the shift vector for the optimum shift magnitude requires at least three calculations of M, i.e. three FFT's plus some additional calculations. Thus the overall computer time

required for a single cycle of refinement is approximately four times that required for one FFT. It is apparent that space group specific FFT's can substantially reduce the required computer time per cycle. Included in the refinement package is a program (to be described elsewhere) that will calculate space group specific FFT's for most non-centrosymmetric space groups.

Crystallographic Term

The function that we have chosen to minimize is

$$M = \sum_{\underline{s}} W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]^2 \qquad (3.3)$$

where F_o and F_c are the observed and calculated structure factors and k is a scale factor. Note that no exponential factor is applied to the F_c's. This will force the thermal factors of the individual atoms to include any "overall" mismatch between the observed and calculated data sets.

At the beginning of each cycle of refinement the scale factor k is determined by minimizing

$$M(k,B) = \sum_{\underline{s}} W(\underline{s})[k \mid F_o(\underline{s}) \mid - \exp(-Bs^2/4) \mid F_c(\underline{s},\underline{p}) \mid]^2 \qquad (3.4)$$

where $s = \sin\theta/\lambda$ and $F_c(\underline{s})$ is treated as a constant. Equation (3.4) includes an overall thermal factor B, which is necessary to allow for an initial overall discrepancy between the F_o's and F_c's. Although both k and B are treated as variables in the minimization of Equation (3.4),

only k is substituted in Equation (3.3). As the refinement proceeds, the discrepancy represented by B is absorbed within the thermal factors of the individual atoms and, during successive cycles, rapidly approaches zero.

Each module of the refinement package is able to calculate both the value and the gradient of its term. For the crystallographic term the structure factors are calculated by a space group specific FFT (Ten Eyck, 1977). The gradients are calculated by a modified version of the procedure outlined by Agarwal (1978). This modified version was devised by A. Lifchitz (see Isaacs, 1982) and is described in detail in Appendix A. In outline, the procedure is as follows. An (F_o-F_c) map is calculated for the molecular volume. For each parameter in the model a convolution, evaluated at the atomic position, is calculated between this map and the derivative of the calculated atomic electron density function for the atom involved. Because the extent of each atom's electron cloud is small, the calculation of this convolution is rapid. Usually the calculation of the convolutions takes about a quarter of the time required to calculate the difference map.

Stereochemical Term

The major goal in the implementation of this part of the package was to make it as easy as possible for the user to specify "ideal" bond lengths and angles. Stereochemical restraints are usually introduced either as energy terms (e.g. see Jack & Levitt, 1978) or by expressing all types of stereochemical restraints as distances (e.g. see Ten Eyck

et al., 1976; Dodson et al., 1976; Konnert & Hendrickson, 1980). There are drawbacks to both such approaches. In the first case it may be difficult to obtain reliable energy parameters, especially for novel chemical groups. Also the introduction of an inappropriate energy term might mask interesting and unexpected features of the structure. On the other hand, if one attempts to define standard geometry in terms of interatomic distances, then such distances must be determined indirectly from a known example with ideal geometry. There are obvious difficulties if no known structure exists for the chemical group in question.

The method used in this package is to include stereochemical restraints as "observations" but to specify such restraints in a form that is most convenient for the user, i.e. as bond lengths, bond angles and so on. There are six classes of stereochemical information with which the structural model can be restrained: bond lengths, bond angles, torsion angles, trigonal planarity, general planarity and contacts between non-bonded atoms. (Chirality is monitored but not restrained because the chirality function is discontinuous and has no derivatives.) Because the program deals directly with the stereochemical information, some of the derivatives are difficult to derive, and, for the planarity restraints, certain assumptions were used to simplify the calculation. The derivations of the gradients for the stereochemical term are given in Appendix B.

To apply the stereochemical restraints the contents of the asymmetric unit are broken up into different hierarchical units. Each unit can be broken up into small subgroups of atoms in whatever manner

is appropriate for the problem at hand. For example, consider the crystal structure of cro repressor (Anderson et al., 1981). The asymmetric unit consists of four chemically identical polypeptides, each with 66 amino acids. The first hierarchical unit is defined by CHAIN cards. In this example we specify that there are four chains, named O, A, B and C, each chain being of type "CRO" (see Fig. 2 for representative data cards). The makeup of a "CRO" chain is then defined by RESIDUE cards. A series of such cards is used to define the sequence of units in the chain (in this case, amino acid residues) and the types of linkages between successive units (in this case peptide bonds). In this example, the units of the chain are named GLY, ALA, THR,..., etc., and the linkages PEPTIDE, SS,..., etc. The geometric restraints associated with each unit or linkage type are defined with GEOMETRY cards (Fig. 2). Each restraint (bond length, bond angle, torsion angle, plane...) is specified in a straightforward manner. There is no particular order in which these cards must be given and they can be arranged into different files in any desired manner.

The enumeration of all the stereochemical restraints in this manner may seem to be time consuming, but most of the files, once created, can be transferred from one application to another. Also it is easy to inspect and alter the ideal values of the restraints since they appear in the program in the same form as in everyday usage. Table 1 gives the library of "ideal" stereochemistry that has been adopted in this laboratory. The individual values come from several sources including Bowen et al. (1958) and Vijayan (1976).

First in the example comes the CHAIN cards which define that all four chains: O, A, B, and C are of type CRO. Next come the cards (not all of which are shown) which define the meaning of type CRO. A type is given for each amino acid in CRO along with the targets and types of any linkages between this residue and other residues. Disulfide bonds are defined in the same manner as peptide bonds. Next the residue types and linkage types are defined by supplying the restraints associated with each. In the example only the definitions of methionine and the linkage for a peptide bond are shown. The restraints for a linkage often involve atoms of both of the involved residues. The atoms of these two classes must be distinguished. This is done by putting a minus sign in front of the atom name for atoms from the amino acid mentioned first on the RESIDUE card and a plus ign in the name of the target amino acid.

There is an unusual feature in the definition of the standard values for torsion angles. The thousand's digit of the standard value in the number of equally spaced minima in the function. The three least significant digits, along with the sign, form the location (in degrees) of one of the minima. The locations of the other minima can be deduced from this information.

```
                    CHAIN O CRO
                    CHAIN A CRO
                    CHAIN B CRO
                    CHAIN C CRO

RESIDUE CRO | 1 MET    2 PEPTIDE
RESIDUE CRO | 2 GLU    3 PEPTIDE
RESIDUE CRO | 3 GLN    4 PEPTIDE
RESIDUE CRO | 4 ARG    5 PEPTIDE
RESIDUE CRO | 5 ILE    6 PEPTIDE
      .
      .
      .

GEOMETRY MET       BOND     1.54    0.02   CA, CB
GEOMETRY MET       BOND     1.52    0.02   CB, CG
GEOMETRY MET       BOND     1.81    0.02   CG, SD
GEOMETRY MET       BOND     1.81    0.02   SD, CE
GEOMETRY MET       ANGLE    112     3      N, CA, CB
GEOMETRY MET       ANGLE    111     3      C, CA, CB
GEOMETRY MET       ANGLE    113     3      CA, CB, CG
GEOMETRY MET       ANGLE    113     3      CB, CG, SD
GEOMETRY MET       ANGLE    100.4   3      CG, SD, CE
GEOMETRY MET       TORS     3060    15     N, CA, CB, CG
GEOMETRY MET       TORS     3060    15     CA, CB, CG, SD
GEOMETRY MET       TORS     3060    15     CB, CG, SD, CE
GEOMETRY MET       CHIRAL   1       1      CA, N, CB, C

GEOMETRY PEPTIDE   BOND     1.45    0.02   -N, -CA
GEOMETRY PEPTIDE   BOND     1.52    0.02   -CA, -C
GEOMETRY PEPTIDE   BOND     1.33    0.02   -C, +N
GEOMETRY PEPTIDE   BOND     1.23    0.02   -C, -O
GEOMETRY PEPTIDE   ANGLE    112     3      -N, -CA, -C
GEOMETRY PEPTIDE   ANGLE    121.1   3      -CA, -C, -O
GEOMETRY PEPTIDE   ANGLE    115.6   3      -CA, -C, +N
GEOMETRY PEPTIDE   ANGLE    123.2   3      -O, -C, +N
GEOMETRY PEPTIDE   ANGLE    121.9   3      -C, +N, +CA
GEOMETRY PEPTIDE   PLANE    5       0.02   -C, -CA, -O, +N, +CA
GEOMETRY PEPTIDE   TORS     2160    30     -N, -CA, -C, +N
GEOMETRY PEPTIDE   TORS     2180    10     -CA, -C, +N, +CA
GEOMETRY PEPTIDE   TORS     3060    20     -C, +N, +CA, +C
```

FIGURE 2. A Structure Definition Example

TABLE 1

Geometry Restraint Library

This table lists the geometry restraint library used when refining a protein structure. It is listed exactly in the form read by the programs of the package (except one card has been broken into two lines to fit the page). The order of the cards is unimportant. However, they have been clustered such that all restraints for a particular unit are next to each other. The first set of restraints define the geometry of a peptide bond. To make the definition of the twenty individual amino acids easier, the restraints common to all amino acids are included in the peptide bond definition. The second group of restraints (CTERM) defines the restraints for the C-terminus of a peptide chain. The next group (BREAK) defines the geometry of an arbitrary break in the main chain of the peptide. This is used in those cases where the electron density is so weak that the peptide model cannot be continued. BREAK is defined as a linkage so that it can be used regardless of the amino acid type of the last residue before the break in the model. SS defines the geometry of a dilsulfide bond. Then follows the definition of the restraints in the side chains of the amino acids.

The format of the cards is described in Figure 2. The values used in the restraints were primarily derived from "The Handbook of Biochemistry and Molecular Biology" (CRC Press, 1976). Because this method of defining restraints is very easy to use, the standard values can be changed very quickly when better values become available.

TABLE 1. Geometry Restraint Library

GEOMETRY	PEPTIDE	BOND	1.45	0.02	N, CA
GEOMETRY	PEPTIDE	BOND	1.52	0.02	CA, C
GEOMETRY	PEPTIDE	BOND	1.33	0.02	C, +N
GEOMETRY	PEPTIDE	BOND	1.23	0.02	C, O
GEOMETRY	PEPTIDE	ANGLE	112	3	N, CA, C
GEOMETRY	PEPTIDE	ANGLE	121.1	3	CA, C, O
GEOMETRY	PEPTIDE	ANGLE	115.6	3	CA, C, +N
GEOMETRY	PEPTIDE	ANGLE	123.2	3	O, C, +N
GEOMETRY	PEPTIDE	ANGLE	121.9	3	C, +N, +CA
GEOMETRY	PEPTIDE	PLANE	5	0.02	C, CA, O, +N, +CA
GEOMETRY	PEPTIDE	TORS	2160	30	N, CA, C, +N
GEOMETRY	PEPTIDE	TORS	2180	10	CA, C, +N, +CA
GEOMETRY	PEPTIDE	TORS	3060	20	C, +N, +CA, +C
GEOMETRY	CTERM	BOND	1.45	0.02	N, CA
GEOMETRY	CTERM	BOND	1.52	0.02	CA, C
GEOMETRY	CTERM	BOND	1.25	0.02	C, O
GEOMETRY	CTERM	BOND	1.25	0.02	C, +OH
GEOMETRY	CTERM	ANGLE	118	3	CA, C, O
GEOMETRY	CTERM	ANGLE	118	3	CA, C, +OH
GEOMETRY	CTERM	ANGLE	112	3	N, CA, C
GEOMETRY	CTERM	TRIG	0	0.02	C, CA, O, +OH
GEOMETRY	BREAK	BOND	1.45	0.02	N, CA
GEOMETRY	BREAK	BOND	1.52	0.02	CA, C
GEOMETRY	BREAK	BOND	1.23	0.02	C, O
GEOMETRY	BREAK	ANGLE	121.2	3	CA, C, O
GEOMETRY	BREAK	ANGLE	110	3	N, CA, C
GEOMETRY	SS	BOND	2.036	0.03	SG, +SG
GEOMETRY	SS	ANGLE	104.5	3	CB, SG, +SG
GEOMETRY	SS	ANGLE	104.5	3	+CB, +SG, SG
GEOMETRY	SS	TORS	2090	10	CB, SG, +SG, +CB
GEOMETRY	ALA	BOND	1.52	0.02	CA, CB
GEOMETRY	ALA	ANGLE	110.9	3	N, CA, CB
GEOMETRY	ALA	ANGLE	111	3	C, CA, CB
GEOMETRY	ALA	CHIRAL	1	1	CA, N, CB, C
GEOMETRY	SER	BOND	1.52	0.02	CA, CB
GEOMETRY	SER	BOND	1.414	0.02	CB, OG
GEOMETRY	SER	ANGLE	112	3	N, CA, CB
GEOMETRY	SER	ANGLE	111	3	C, CA, CB
GEOMETRY	SER	ANGLE	110	3	CA, CB, OG
GEOMETRY	SER	TORS	3060	15	N, CA, CB, OG
GEOMETRY	SER	CHIRAL	1	1	CA, N, CB, C
GEOMETRY	CYS	BOND	1.54	0.02	CA, CB
GEOMETRY	CYS	BOND	1.81	0.02	CB, SG
GEOMETRY	CYS	ANGLE	112	3	N, CA, CB
GEOMETRY	CYS	ANGLE	111	3	C, CA, CB
GEOMETRY	CYS	ANGLE	113	3	CA, CB, SG
GEOMETRY	CYS	TORS	3060	15	N, CA, CB, SG

TABLE 1. (continued)

GEOMETRY CYS	CHIRAL	1	1	CA, N, CB, C
GEOMETRY MET	BOND	1.54	0.02	CA, CB
GEOMETRY MET	BOND	1.52	0.02	CB, CG
GEOMETRY MET	BOND	1.81	0.02	CG, SD
GEOMETRY MET	BOND	1.81	0.02	SD, CE
GEOMETRY MET	ANGLE	112	3	N, CA, CB
GEOMETRY MET	ANGLE	111	3	C, CA, CB
GEOMETRY MET	ANGLE	113	3	CA, CB, CG
GEOMETRY MET	ANGLE	113	3	CB, CG, SD
GEOMETRY MET	ANGLE	100.4	3	CG, SD, CE
GEOMETRY MET	TORS	3060	15	N, CA, CB, CG
GEOMETRY MET	TORS	3060	15	CA, CB, CG, SD
GEOMETRY MET	TORS	3060	15	CB, CG, SD, CE
GEOMETRY MET	CHIRAL	1	1	CA, N, CB, C
GEOMETRY LYS	BOND	1.52	0.02	CA, CB
GEOMETRY LYS	BOND	1.52	0.02	CB, CG
GEOMETRY LYS	BOND	1.52	0.02	CG, CD
GEOMETRY LYS	BOND	1.52	0.02	CD, CE
GEOMETRY LYS	BOND	1.47	0.02	CE, NZ
GEOMETRY LYS	ANGLE	112	3	N, CA, CB
GEOMETRY LYS	ANGLE	111	3	C, CA, CB
GEOMETRY LYS	ANGLE	113	3	CA, CB, CG
GEOMETRY LYS	ANGLE	113	3	CB, CG, CD
GEOMETRY LYS	ANGLE	113	3	CG, CD, CE
GEOMETRY LYS	ANGLE	113	3	CD, CE, NZ
GEOMETRY LYS	TORS	3060	15	N, CA, CB, CG
GEOMETRY LYS	TORS	3060	15	CA, CB, CG, CD
GEOMETRY LYS	TORS	3060	15	CB, CG, CD, CE
GEOMETRY LYS	TORS	3060	15	CG, CD, CE, NZ
GEOMETRY LYS	CHIRAL	1	1	CA, N, CB, C
GEOMETRY VAL	BOND	1.524	0.02	CA, CB
GEOMETRY VAL	BOND	1.524	0.02	CB, CG1
GEOMETRY VAL	BOND	1.524	0.02	CB, CG2
GEOMETRY VAL	ANGLE	112	3	N, CA, CB
GEOMETRY VAL	ANGLE	111	3	C, CA, CB
GEOMETRY VAL	ANGLE	110	3	CA, CB, CG1
GEOMETRY VAL	ANGLE	110	3	CA, CB, CG2
GEOMETRY VAL	ANGLE	110	3	CG1, CB, CG2
GEOMETRY VAL	TORS	3060	15	N, CA, CB, CG1
GEOMETRY VAL	CHIRAL	1	1	CA, N, CB, C
GEOMETRY THR	BOND	1.54	0.02	CA, CB
GEOMETRY THR	BOND	1.414	0.02	CB, OG1
GEOMETRY THR	BOND	1.524	0.02	CB, CG2
GEOMETRY THR	ANGLE	112	3	N, CA, CB
GEOMETRY THR	ANGLE	111	3	C, CA, CB
GEOMETRY THR	ANGLE	110	3	CA, CB, OG1
GEOMETRY THR	ANGLE	110	3	CA, CB, CG2

TABLE 1. (continued)

GEOMETRY THR	ANGLE	110	3	OG1, CB, CG2
GEOMETRY THR	TORS	3060	15	N, CA, CB, OG1
GEOMETRY THR	CHIRAL	1	1	CA, N, CB, C
GEOMETRY THR	CHIRAL	1	1	CB, CA, CG2, OG1
GEOMETRY ILE	BOND	1.524	0.02	CA, CB
GEOMETRY ILE	BOND	1.524	0.02	CB, CG1
GEOMETRY ILE	BOND	1.524	0.02	CG1, CD1
GEOMETRY ILE	BOND	1.524	0.02	CB, CG2
GEOMETRY ILE	ANGLE	112	3	N, CA, CB
GEOMETRY ILE	ANGLE	111	3	C, CA, CB
GEOMETRY ILE	ANGLE	110	3	CA, CB, CG2
GEOMETRY ILE	ANGLE	110	3	CA, CB, CG1
GEOMETRY ILE	ANGLE	110	3	CG1, CB, CG2
GEOMETRY ILE	ANGLE	113	3	CB, CG1, CD1
GEOMETRY ILE	TORS	3060	15	N, CA, CB, CG1
GEOMETRY ILE	TORS	3060	15	CA, CB, CG1, CD1
GEOMETRY ILE	CHIRAL	1	1	CA, N, CB, C
GEOMETRY ILE	CHIRAL	1	1	CB, CA, CG2, CG1
GEOMETRY LEU	BOND	1.54	0.02	CA, CB
GEOMETRY LEU	BOND	1.524	0.02	CB, CG
GEOMETRY LEU	BOND	1.524	0.02	CG, CD1
GEOMETRY LEU	BOND	1.524	0.02	CG, CD2
GEOMETRY LEU	ANGLE	112	3	N, CA, CB
GEOMETRY LEU	ANGLE	111	3	C, CA, CB
GEOMETRY LEU	ANGLE	113	3	CA, CB, CG
GEOMETRY LEU	ANGLE	110	3	CB, CG, CD1
GEOMETRY LEU	ANGLE	110	3	CB, CG, CD2
GEOMETRY LEU	ANGLE	110	3	CD1, CG, CD2
GEOMETRY LEU	TORS	3060	15	N, CA, CB, CG
GEOMETRY LEU	TORS	3060	15	CA, CB, CG, CD1
GEOMETRY LEU	CHIRAL	1	1	CA, N, CB, C
GEOMETRY ASP	BOND	1.54	0.02	CA, CB
GEOMETRY ASP	BOND	1.502	0.02	CB, CG
GEOMETRY ASP	BOND	1.217	0.02	CG, OD1
GEOMETRY ASP	BOND	1.310	0.02	CG, OD2
GEOMETRY ASP	ANGLE	112	3	N, CA, CB
GEOMETRY ASP	ANGLE	111	3	C, CA, CB
GEOMETRY ASP	ANGLE	113	3	CA, CB, CG
GEOMETRY ASP	ANGLE	123	3	CB, CG, OD1
GEOMETRY ASP	ANGLE	112.6	3	CB, CG, OD2
GEOMETRY ASP	ANGLE	123.6	3	OD1, CG, OD2
GEOMETRY ASP	TRIG	0.0	0.02	CG, CB, OD1, OD2
GEOMETRY ASP	TORS	3060	15	N, CA, CB, CG
GEOMETRY ASP	TORS	2165	30	CA, CB, CG, OD1
GEOMETRY ASP	CHIRAL	1	1	CA, N, CB, C
GEOMETRY ASN	BOND	1.54	0.02	CA, CB
GEOMETRY ASN	BOND	1.516	0.02	CB, CG

TABLE 1. (continued)

GEOMETRY ASN	BOND	1.24	0.02	CG, OD1
GEOMETRY ASN	BOND	1.33	0.02	CG, ND2
GEOMETRY ASN	ANGLE	112	3	N, CA, CB
GEOMETRY ASN	ANGLE	111	3	C, CA, CB
GEOMETRY ASN	ANGLE	113	3	CA, CB, CG
GEOMETRY ASN	ANGLE	121	3	CB, CG, OD1
GEOMETRY ASN	ANGLE	116	3	CB, CG, ND2
GEOMETRY ASN	ANGLE	123	3	OD1, CG, ND2
GEOMETRY ASN	TRIG	0	0.02	CG, CG, OD1, ND2
GEOMETRY ASN	TORS	3060	15	N, CA, CB, CG
GEOMETRY ASN	TORS	2165	30	CA, CB, CG, OD1
GEOMETRY ASN	CHIRAL	1	1	CA, N, CB, C
GEOMETRY GLU	BOND	1.54	0.02	CA, CB
GEOMETRY GLU	BOND	1.54	0.02	CB, CG
GEOMETRY GLU	BOND	1.502	0.02	CG, CD
GEOMETRY GLU	BOND	1.217	0.02	CD, OE1
GEOMETRY GLU	BOND	1.310	0.02	CD, OE2
GEOMETRY GLU	ANGLE	112	3	N, CA, CB
GEOMETRY GLU	ANGLE	111	3	C, CA, CB
GEOMETRY GLU	ANGLE	113	3	CA, CB, CG
GEOMETRY GLU	ANGLE	113	3	CB, CG, CD
GEOMETRY GLU	ANGLE	123.0	3	CG, CD, OE1
GEOMETRY GLU	ANGLE	112.6	3	CG, CD, OE2
GEOMETRY GLU	ANGLE	123.6	3	OE1, CD, OE2
GEOMETRY GLU	TRIG	0	0.02	CD, CG, OE1, OE2
GEOMETRY GLU	TORS	3060	15	N, CA, CB, CG
GEOMETRY GLU	TORS	3060	15	CA, CB, CG, CD
GEOMETRY GLU	TORS	2165	30	CB, CG, CD, OE1
GEOMETRY GLU	CHIRAL	1	1	CA, N, CB, C
GEOMETRY GLN	BOND	1.54	0.02	CA, CB
GEOMETRY GLN	BOND	1.54	0.02	CB, CG
GEOMETRY GLN	BOND	1.516	0.02	CG, CD
GEOMETRY GLN	BOND	1.24	0.02	CD, OE1
GEOMETRY GLN	BOND	1.33	0.02	CD, NE2
GEOMETRY GLN	ANGLE	112	3	N, CA, CB
GEOMETRY GLN	ANGLE	111	3	C, CA, CB
GEOMETRY GLN	ANGLE	113	3	CA, CB, CG
GEOMETRY GLN	ANGLE	113	3	CB, CG, CD
GEOMETRY GLN	ANGLE	121	3	CG, CD, OE1
GEOMETRY GLN	ANGLE	116	3	CG, CD, NE2
GEOMETRY GLN	ANGLE	123	3	OE1, CD, NE2
GEOMETRY GLN	TRIG	0	0.02	CD, CG, OE1, NE2
GEOMETRY GLN	TORS	3060	15	N, CA, CB, CG
GEOMETRY GLN	TORS	3060	15	CA, CB, CG, CD
GEOMETRY GLN	TORS	2165	30	CB, CG, CD, OE1
GEOMETRY GLN	CHIRAL	1	1	CA, N, CB, C
GEOMETRY ARG	BOND	1.524	0.02	CA, CB

TABLE 1. (continued)

GEOMETRY ARG	BOND	1.524	0.02	CB, CG
GEOMETRY ARG	BOND	1.524	0.02	CG, CD
GEOMETRY ARG	BOND	1.46	0.02	CD, NE
GEOMETRY ARG	BOND	1.33	0.02	NE, CZ
GEOMETRY ARG	BOND	1.33	0.02	CZ, NH1
GEOMETRY ARG	BOND	1.33	0.02	CZ, NH2
GEOMETRY ARG	ANGLE	112	3	N, CA, CB
GEOMETRY ARG	ANGLE	111	3	C, CA, CB
GEOMETRY ARG	ANGLE	113	3	CA, CB, CG
GEOMETRY ARG	ANGLE	113	3	CB, CG, CD
GEOMETRY ARG	ANGLE	113	3	CG, CD, NE
GEOMETRY ARG	ANGLE	124	3	CD, NE, CZ
GEOMETRY ARG	ANGLE	122	3	NE, CZ, NH1
GEOMETRY ARG	ANGLE	119	3	NE, CZ, NH2
GEOMETRY ARG	ANGLE	119	3	NH1, CZ, NH2
GEOMETRY ARG	PLANE	5	0.02	CD, NE, CZ, NH1, NH2
GEOMETRY ARG	TORS	3060	15	N, CA, CB, CG
GEOMETRY ARG	TORS	3060	15	CA, CB, CG, CD
GEOMETRY ARG	TORS	3060	15	CB, CG, CD, NE
GEOMETRY ARG	TORS	4000	15	CG, CD, NE, CZ
GEOMETRY ARG	TORS	2000	10	CD, NE, CZ, NH1
GEOMETRY ARG	CHIRAL	1	1	CA, N, CB, C
GEOMETRY PRO	BOND	1.53	0.02	CA, CB
GEOMETRY PRO	BOND	1.52	0.02	CB, CG
GEOMETRY PRO	BOND	1.52	0.02	CG, CD
GEOMETRY PRO	BOND	1.49	0.02	CD, N
GEOMETRY PRO	ANGLE	110	3	N, CA, CB
GEOMETRY PRO	ANGLE	110	3	C, CA, CB
GEOMETRY PRO	ANGLE	110	3	CA, CB, CG
GEOMETRY PRO	ANGLE	110	3	CB, CG, CD
GEOMETRY PRO	ANGLE	110	3	CG, CD, N
GEOMETRY PRO	ANGLE	116	3	CD, N, CA
GEOMETRY PRO	CHIRAL	1	1	CA, N, CB, C
GEOMETRY HIS	BOND	1.54	0.02	CA, CB
GEOMETRY HIS	BOND	1.495	0.02	CB, CG
GEOMETRY HIS	BOND	1.385	0.02	CG, ND1
GEOMETRY HIS	BOND	1.327	0.02	ND1, CE1
GEOMETRY HIS	BOND	1.331	0.02	CE1, NE2
GEOMETRY HIS	BOND	1.376	0.02	NE2, CD2
GEOMETRY HIS	BOND	1.360	0.02	CG, CD2
GEOMETRY HIS	ANGLE	112	3	N, CA, CB
GEOMETRY HIS	ANGLE	111	3	C, CA, CB
GEOMETRY HIS	ANGLE	113	3	CA, CB, CG
GEOMETRY HIS	ANGLE	130.2	3	CB, CG, CD2
GEOMETRY HIS	ANGLE	121.5	3	CB, CG, ND1
GEOMETRY HIS	ANGLE	107.7	3	ND1, CG, CD2
GEOMETRY HIS	ANGLE	106.5	3	CG, ND1, CE1

TABLE 1. (continued)

GEOMETRY HIS	ANGLE	110.1	3	ND1, CE1, NE2
GEOMETRY HIS	ANGLE	108.5	3	CE1, NE2, CD2
GEOMETRY HIS	ANGLE	106.5	3	CG, CD2, NE2
GEOMETRY HIS	PLANE	6	0.02	CB, CG, ND1, CE1, CD2, NE2
GEOMETRY HIS	TORS	3060	15	N, CA, CB, CG
GEOMETRY HIS	TORS	2090	20	CA, CB, CG, ND1
GEOMETRY HIS	CHIRAL	1	1	CA, N, CB, C
GEOMETRY PHE	BOND	1.54	0.02	CA, CB
GEOMETRY PHE	BOND	1.51	0.02	CB, CG
GEOMETRY PHE	BOND	1.395	0.01	CG, CD1
GEOMETRY PHE	BOND	1.395	0.01	CG, CD2
GEOMETRY PHE	BOND	1.395	0.01	CD1, CE1
GEOMETRY PHE	BOND	1.395	0.01	CD2, CE2
GEOMETRY PHE	BOND	1.395	0.01	CZ, CE1
GEOMETRY PHE	BOND	1.395	0.01	CZ, CE2
GEOMETRY PHE	ANGLE	112	3	N, CA, CB
GEOMETRY PHE	ANGLE	111	3	C, CA, CB
GEOMETRY PHE	ANGLE	113	3	CA, CB, CG
GEOMETRY PHE	ANGLE	120.7	3	CB, CG, CD1
GEOMETRY PHE	ANGLE	120.7	3	CB, CG, CD2
GEOMETRY PHE	ANGLE	118.2	3	CD1, CG, CD2
GEOMETRY PHE	ANGLE	120.3	3	CE1, CZ, CE2
GEOMETRY PHE	ANGLE	121.2	3	CG, CD1, CE1
GEOMETRY PHE	ANGLE	121.2	3	CG, CD2, CE2
GEOMETRY PHE	ANGLE	119.5	3	CZ, CE1, CD1
GEOMETRY PHE	ANGLE	119.5	3	CZ, CE2, CD2
GEOMETRY PHE	PLANE	7	0.02	CB,CG,CD1,CE1,CZ,CE2,CD2
GEOMETRY PHE	TORS	3060	15	N, CA, CB, CG
GEOMETRY PHE	TORS	2090	20	CA, CB, CG, CD1
GEOMETRY PHE	CHIRAL	1	1	CA, N, CB, C
GEOMETRY TYR	BOND	1.54	0.02	CA, CB
GEOMETRY TYR	BOND	1.51	0.02	CB, CG
GEOMETRY TYR	BOND	1.395	0.01	CG, CD1
GEOMETRY TYR	BOND	1.395	0.01	CG, CD2
GEOMETRY TYR	BOND	1.395	0.01	CD1, CE1
GEOMETRY TYR	BOND	1.395	0.01	CD2, CE2
GEOMETRY TYR	BOND	1.395	0.01	CZ, CE1
GEOMETRY TYR	BOND	1.395	0.01	CZ, CE2
GEOMETRY TYR	BOND	1.37	0.02	CZ, OH
GEOMETRY TYR	ANGLE	112	3	N, CA, CB
GEOMETRY TYR	ANGLE	111	3	C, CA, CB
GEOMETRY TYR	ANGLE	113	3	CA, CB, CG
GEOMETRY TYR	ANGLE	120.7	3	CB, CG, CD1
GEOMETRY TYR	ANGLE	120.7	3	CB, CG, CD2
GEOMETRY TYR	ANGLE	118.2	3	CD1, CG, CD2
GEOMETRY TYR	ANGLE	120.3	3	CE1, CZ, CE2
GEOMETRY TYR	ANGLE	121.2	3	CG, CD1, CE1

TABLE 1. (continued)

GEOMETRY TYR	ANGLE	121.2	3	CG, CD2, CE2
GEOMETRY TYR	ANGLE	119.5	3	CZ, CE1, CD1
GEOMETRY TYR	ANGLE	119.5	3	CZ, CE2, CD2
GEOMETRY TYR	ANGLE	120	3	CE1, CZ, OH
GEOMETRY TYR	ANGLE	120	3	CE2, CZ, OH
GEOMETRY TYR	PLANE	8	0.02	CB,CG,CD1,CE1,CZ,CE2,CD2,OH
GEOMETRY TYR	TORS	3060	15	N, CA, CB, CG
GEOMETRY TYR	TORS	2090	20	CA, CB, CG, CD1
GEOMETRY TYR	CHIRAL	1	1	CA, N, CB, C
GEOMETRY TRP	BOND	1.54	0.02	CA, CB
GEOMETRY TRP	BOND	1.497	0.02	CB, CG
GEOMETRY TRP	BOND	1.355	0.02	CG, CD1
GEOMETRY TRP	BOND	1.368	0.02	CD1, NE1
GEOMETRY TRP	BOND	1.368	0.02	NE1, CE2
GEOMETRY TRP	BOND	1.395	0.02	CE2, CZ2
GEOMETRY TRP	BOND	1.376	0.02	CZ2, CH2
GEOMETRY TRP	BOND	1.395	0.02	CH2, CZ3
GEOMETRY TRP	BOND	1.376	0.02	CZ3, CE3
GEOMETRY TRP	BOND	1.392	0.02	CD2, CE3
GEOMETRY TRP	BOND	1.433	0.02	CG, CD2
GEOMETRY TRP	BOND	1.409	0.02	CD2, CE2
GEOMETRY TRP	ANGLE	112	3	N, CA, CB
GEOMETRY TRP	ANGLE	111	3	C, CA, CB
GEOMETRY TRP	ANGLE	113	3	CA, CB, CG
GEOMETRY TRP	ANGLE	126.1	3	CB, CG, CD1
GEOMETRY TRP	ANGLE	127.6	3	CB, CG, CD2
GEOMETRY TRP	ANGLE	106.3	3	CD2, CG, CD1
GEOMETRY TRP	ANGLE	110.4	3	CG, CD1, NE1
GEOMETRY TRP	ANGLE	109.1	3	CD1, NE1, CE2
GEOMETRY TRP	ANGLE	107.2	3	NE1, CE2, CD2
GEOMETRY TRP	ANGLE	107.0	3	CG, CD2, CE2
GEOMETRY TRP	ANGLE	119.0	3	CE2, CD2, CE3
GEOMETRY TRP	ANGLE	121.9	3	CD2, CE2, CZ2
GEOMETRY TRP	ANGLE	117.8	3	CE2, CZ2, CH2
GEOMETRY TRP	ANGLE	121.3	3	CZ2, CH2, CZ3
GEOMETRY TRP	ANGLE	121.1	3	CE3, CZ3, CH2
GEOMETRY TRP	ANGLE	119.0	3	CD2, CE3, CZ3
GEOMETRY TRP	ANGLE	134.0	3	CG, CD2, CE3
GEOMETRY TRP	ANGLE	130.9	3	NE1, CE2, CZ2
GEOMETRY TRP	PLANE	10	0.02	CB,CG,CD1,NE1,CD2,CE2,CZ2, CH2, CE3, CZ3
GEOMETRY TRP	TORS	3060	15	N, CA, CB, CG
GEOMETRY TRP	TORS	2090	20	CA, CB, CG, CD2
GEOMETRY TRP	CHIRAL	1	1	CA, N, CB, C

Interactions between non-bonded atoms cannot be defined in the manner described above because one does not know in advance which atoms may approach each other. Close contacts are discovered by generating a list of all pairs of atoms which are closer to each other than specified values and discarding from consideration any pairs which are bonded, or are involved in a 1-3 or 1-4 type contacts. The 1-3 and 1-4 contacts are better dealt with as bond angle and torsion angles. The standard value for the closest distance allowed before an action is taken is defined in terms of the elemental type of the two atoms. This method of definition allows a closer approach between atoms which have the potential of forming a hydrogen bond, or a salt bridge than the distance allowed for atoms in van der Waals contact. The program will prevent non-bonded atoms from moving too close together but no attractive force is applied to atoms that are beyond the specified approach distance.

One novel feature of the program is the ability to avoid steric clashes between adjacent molecules in the crystal. By specifying the appropriate symmetry operators the list of potential non-bonded contacts can be extended to include molecules that surround the reference structure. This procedure is particularly useful in avoiding "duplicate" or "overlapping" solvent atoms.

The program that implements the stereochemistry module has a number of additional features. It can list the worst discrepancies in the model for each type of geometry restraint and provide overall statistics for each class. Also it can produce a table which compares the "ideal" value of each restraint with the average value in the present model.

This table is useful when looking for potential errors in the geometry library.

The Control Program

The control program has two major functions: (1) to determine the overall direction of shift for each parameter, and (2) to determine the optimum magnitude (fraction of the shift) to be applied along the shift vector.

In an initial cycle of refinement, the overall gradient is obtained by combining the contributions from the crystallographic, stereochemical and any other terms and the direction of shift is obtained by the method of steepest descent. For second and subsequent cycles of refinement one can combine the overall gradient with the direction of shift used in the previous cycle and determine the new shift direction by the conjugate gradient technique. The control program also determines the optimal shift magnitude by searching along the shift direction as described previously.

A number of options exist for modifying the shift vector before the shift is applied. One can combine all or part of the structure into units within which all the atoms are treated identically (e.g. see Fig. 3). If the atoms within an amino acid are combined, the temperature factor shift applied to each atom will be the average of the individual shifts. Positional parameters can be treated as though the group were a rigid body. This is done by fitting an overall rotation and translation to the individual shifts of the atoms by a least squares fitting

(a)

```
COMBINE XYZ 360 - 361
COMBINE XYZ 362 - 363
COMBINE XYZ 364 - 365
COMBINE XYZ 366 - 367
COMBINE XYZ 368 - 369
COMBINE XYZ 370 - 371
COMBINE XYZ 372 - 373
```

(b)

```
FIX OCC 1 - COOH
FIX OCC SOL1
FIX OCC ZINC:ZINC
FIX OCC CAL1:CAL
FIX OCC CAL2:CAL
FIX OCC CAL3:CAL
FIX OCC CAL4:CAL
COMBINE OCC INHIBITOR
FIX B
FIX XYZ
```

FIGURE 3. Several Examples of Parameter Editing Cards

Figure 3a shows the cards required to command the control program to treat the seven Bacteriochlorophyll a molecules as seven rigid bodies. Each molecule contains two residues: the ring and the phytol tail. The first card is read as "combine the positional parameter for all residues between residue 360 and residue 361. The operation "all residues between" is defined as taking the first residue and then marching down the linkage definitions from the RESIDUE cards taking all residues found until the last residue is found.

Figure 3b shows the cards required to refine a single occupancy parameter for an inhibitor of the protein Thermolysin. The first card fixes the occupancy of all of the protein atoms. COOH is the residue name used in this project for the extra oxygen at the C terminus. The next card fixes the occupancy of all of the solvent atoms. They are defined in a single residue to make operations such as this simple. Then the occupancies for the zinc atom and the four calcium atoms are fixed. Finally the occupancy parameters for the inhibitor, presumed to be defined as a single residue called INHIBITOR, are combined. The last two cards simply state that the positional and thermal parameters for all parameters in the model should be held constant.

procedure. This process gives an exact solution in one cycle of minimization. It is also possible to place upper and lower limits on the values of the thermal factors and the occupancies.

Applications

The package of programs has been in productive use (while being developed) for four years. It has been applied to a number of different refinement tasks in this laboratory and elsewhere. In this section we briefly review some of the applications.

Bacteriochlorophyll *a* Protein

The first extensive use of the package was in the refinement of the bacteriochlorophyll *a* protein (Bchl protein) from the photosynthetic bacterium *prosthecochloris aestaurii*. This molecule, molecular weight 150,000 daltons, consists of three identical subunits related by a 3-fold axis of symmetry. Each of the subunits consists of a polypeptide chain of approximately 350 amino acids that encloses seven bacteriochlorophyll *a* molecules (Fenna & Matthews, 1975; Matthews *et al.*, 1979). During the course of the refinement the amino acid sequence of the protein was not known, but is now essentially complete (R.E. Fenna, personal communication).

There were several factors that led to the adoption of the present refinement package. The first was the size of the computational problem. As summarized in Table 2, the asymmetric unit contains 3086 atoms and there are 44,000 reflections to 1.9Å resolution. In addition,

TABLE 2. Representative Refinements of Macromolecular Structures

Protein	Bchl protein	Thermolysin[a]	Thermolysin[b]	T4 lysozyme[c]
Space group	$P6_3$	$P6_122$	$P6_122$	$P3_221$
Cell dimensions				
a, b (Å)	111.9	94.1	94.1	61.2
c (Å)	98.6	131.4	131.4	97.4
Number of atoms	3086	2637	2643	1429
Resolution (Å)	1.9	2.3	1.6	1.7
Number of reflections	44,000	13,523	31,627	15,791
Final R-value	18.9%	17.4%	17.7%	17.4%
Deviations from ideal values:				
Bond lengths (Å)	0.02	0.02	0.02	0.02
Bond angles	3.2	2.9	2.8	2.6
Time per cycle (CPU hours)	2.5	.8	1.5	0.75

[a] Thermolysin:phosphoramidon inhibitor complex (Tronrud et al., 1986).

[b] Complex of thermolysin with carbobenzoxy-Gly^P(OH)-L-Leu-L-Leu (Holden et al., in preparation).

[c] Bacteriophage T4 lysozyme mutant Thr 157 → Glu (unpublished results of T. Alber, K. Wilson, B.W.M et al.).

the space group ($P6_3$) precludes full use of the crystallographic symmetry to reduce the size of the FFT calculations. Experience with other projects in the laboratory suggested that the Hendrickson refinement program (Hendrickson & Konnert, 1980) would require about 48 hrs of c.p.u. time per refinement cycle on our VAX 11/780 (subsequent improvements to the Hendrickson program have substantially improved its computational efficiency). We had also had experience in the laboratory with EREF (Jack & Levitt, 1978). This program is substantially faster, on a per cycle basis, than the Hendrickson program. A disadvantage of using EREF for the refinement of the Bchl protein arose from the presence of the Bchl rings. Because of the diversity of bond lengths and bond angles in the seven bacteriochlorophylls, as well as uncertainties in their energetics, the definition of standard geometry for use by EREF appeared to be quite difficult.

As discussed previously, the definition of standard geometry in the present refinement package is very flexible, and readily adaptable to "unusual" situations. In particular, in the present situation, it was not necessary to force all the conjugated atoms in the Bchl rings to lie in a single plane. Rather, we divided the conjugated atoms into appropriate sets of overlapping sub-planes (see Chapter IV). This method of restraint maintained local planarity, but allowed larger scale deformations. This procedure led to the finding that the seven Bchls exhibit two distinct classes of bending, one of which is also observed in the structure of ethyl chlorophillide _a_ (see Chapter IV). An inappropriate application of restraints to the Bchl rings could well have masked this small but significant effect.

In the initial refinement of the Bchl protein, the refinement package used analytical summations to calculate the gradient of the crystallographic term. With this method, each cycle of refinement required 8 hours of CPU time on our VAX 11/780. The present version of the program requires 2.5 hrs for the same problem.

The general strategy of refinement that we have adopted for the Bchl and other proteins is to first refine for several cycles with weak weights on geometry, then to run a few cycles of temperature factor refinement, then to restore the model to good stereochemistry by refining for several cycles with strong weights and to finish with several additional cycles of thermal factor refinement. At this stage the resulting difference electron density maps and "$2F_o - F_c$" maps are inspected on the graphics system in the usual way. Potential problem areas are highlighted by inspecting the lists of worst bond lengths, bond angles, departures from planarity, thermal factors and largest derivatives of the crystallographic term. In the case of the Bchl protein it was also necessary to consider possible errors in the assumed amino acid sequence. In this instance, the cycles of refinement followed by inspection of the model were repeated seven times to achieve the final refined structure and "X-ray" amino acid sequence. The overall refinement statistics are summarized in Table 2.

Thermolysin-Inhibitor Complexes

The refinement package has been used extensively to study a number of complexes of inhibitors with the thermostable endopeptidase

thermolysin (e.g. see Holmes et al., 1983; see also Chapters 5 and 6). In different instances, the resolution ranged from 2.3Å to 1.6Å.

Because of the hexagonal space group and the size of the problem (Table 2), calculation of structure factors and crystallographic derivatives by conventional methods is time consuming. (Refinement of the native structure at 1.6Å resolution by the method of Hendrickson and Konnert (1980) required 21 CPU hours per cycle (Holmes & Matthews, 1982).) The present program package requires 0.7 to 1.5 hours per cycle, depending on the resolution. The ability of the present program package to specify the geometry of chemically unusual inhibitors is an advantage. In addition, it is also possible to define appropriate stereochemistry for inhibitors that are covalently bonded to the enzyme (e.g. see Holmes et al., 1983).

The refinement of an inhibitor complex normally requires 10-30 cycles of refinement. Little manual intervention is required because thermolysin normally does not change its conformation very much when inhibitors are bound. It is, of course, always necessary to monitor the configuration of the inhibitor during refinement and to check for changes in solvent structure concomitant with inhibitor binding.

Bacteriophage T4 Mutant Structures

As part of a program to determine the roles of individual amino acids in stabilizing protein structures, the structures of a series of mutant T4 phage lysozymes have been determined (e.g. see Grütter et al, 1983; Alber et al., 1986). The refinement of each mutant (unpublished

results) is, in principle, very similar to the refinement of an enzyme-inhibitor complex. One starts with the refined structure of the native protein, locally modified to correspond to the structure of the mutant. In the early stages of the refinement, the stereochemical restraints are kept weak, to allow the starting model to relax to conform to the diffraction data observed for the mutant structure. When there appears to be no systematic shifts in the coordinates from one refinement cycle to the next, the stereochemical restraints are strengthened in order the enforce the "ideal" geometry. Some representative refinement statistics are summarized in Table 2. Experience with refining these mutant structures at resolutions in the range 1.9-1.7Å suggests that the refinement procedure is capable of successfully moving both main-chain and side-chain atoms through distances of 1.0Å.

CHAPTER IV

STRUCTURE AND X-RAY AMINO ACID SEQUENCE OF A BACTERIOCHLOROPHYLL *a*-PROTEIN FROM *PROSTHECOCHLORIS AESTUARII* REFINED AT 1.9Å RESOLUTION

Abstract

The structure of the water-soluble bacteriochlorophyll *a*-protein (Bchl protein) from the green photosynthetic bacterium *Prosthecochloris aestuarii* has been refined at 1.9Å resolution to a crystallographic residual of 18.9%. The refinement was carried out without knowledge of the amino acid sequence and has led to an "X-ray sequence". The structure consists of seven Bchl molecules enclosed within a protein "bag" and the refinement supports the general conformation of the molecule described previously (Matthews *et al.*, 1979). The refinement also supports the previous suggestion that the ligands to the seven Bchl magnesiums are, respectively, five histidines, a carbonyl oxygen from the polypeptide backbone of the protein, and a bound water molecule.

The conformations of the seven Bchl head groups are described in detail. In two cases the magnesium atoms are approximately 0.48Å "below" the plane of the conjugated macrocycle while in the other five cases the metals are, on average, 0.48Å "above" the plane. The acetyl ring substituents are more-or-less coplanar with the dihydrophorbin macrocycle, consistent with a resonance Raman study (Lutz *et al.*, 1982).

The conjugated atoms in each of the seven macrocycles have significant departures from strict planarity. These deviations are similar for Bchls 1, 2 and 3 ("Class I") and are also similar for Bchls 4, 5, 6 and 7 ("Class II"). Ethylchlorophillide (Strouse, 1974) also belongs to Class II. The out-of-plane deformations for the Class I and Class II bacteriochlorophylls appear to correspond to two distinct modes of bending or curvature of the dihydrophorbin macrocycle.

Introduction

The light-gathering apparatus of the green photosynthetic bacteria consists of three chlorophyll-containing entities. First, there are the "chlorophyll bodies" which reside beneath the cytoplasmic membrane. These incorporate 95% of the total chlorophyll of the organism and constitute the principal light-gathering pigment. Second, there is the "bacteriochlorophyll a-protein", which includes most of the remaining chlorophyll, and resides between the chlorophyll bodies and the cytoplasmic membrane. Third, there are the reaction centers. The bacteriochlorophyll a-protein is thought to transmit excitation energy to a special pair of chlorophyll molecules in the reaction center where the first step in energy transduction, namely a charge separation occurs.

The bacteriochlorophyll a-protein of Prosthecochloris aestuarii (Bchl-protein), first isolated and crystallized by J. Olson (1971, 1978), has been subject to crystallographic analysis, and the structure determined to a nominal resolution of 2.8Å (Fenna & Matthews, 1975; Matthews et al., 1979). The molecule, molecular weight 150,000 daltons,

consists of three identical subunits related by 120° rotations about a 3-fold axis of symmetry. In each subunit the polypeptide backbone forms a large, twisted β-sheet of 16 strands that forms the "outside" of the protein molecule (i.e. the part exposed to solvent), and encloses a central core of 7 bacteriochlorophyll a molecules (Fig. 4). The available evidence suggests that the trimeric structure seen in the crystals occurs in vivo and is not an artifact of crystallization.

Several different aspects of this structure are of interest. The interactions between the phytyl chains of the bacteriochlorophyll a molecules and the amino acid side chains provide a good model for lipid-protein interactions. Also, the packing of the bacteriochlorophyll a molecules gives insight into other protein-chlorophyll interactions. In particular, it will be of interest to compare the chlorophyll-chlorophyll and chlorophyll-protein interactions in the Bchl-protein with the recently determined structure of the reaction center from Rhodopseudomonas viridis (Deisenhofer et al., 1984). Theoretical calculations of the absorption and circular dichroism spectra of the bacteriochlorophyll core of this protein have been attempted but the agreement between the observed and calculated spectra is poor (Perlstein & Heminger, 1978).

In this paper we describe the refinement of the Bchl-protein structure at a resolution of 1.9Å. The refinement was carried out without knowledge of the amino acid sequence and has led to an "X-ray sequence". The "X-ray sequence" at an intermediate stage of refinement has been reported by Schmid et al. (1983). In this paper we report the

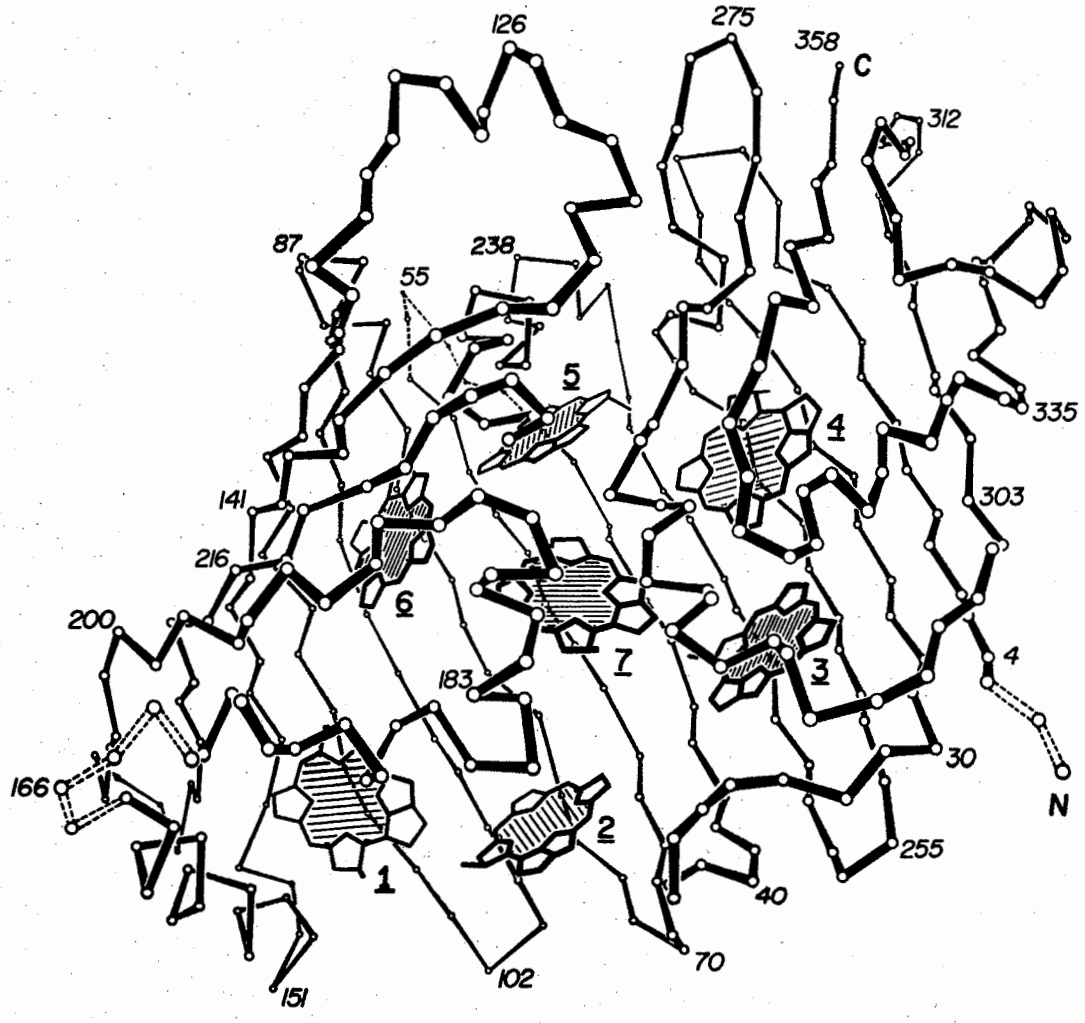


FIGURE 4. Alpha-carbon Backbone of a Bchl-Protein Monomer

Simplified view of one subunit of the Bchl-protein, viewed from the center of the trimer, with the three-fold axis horizontal. For clarity the phytyl chains and other bacteriochlorophyll ring substituents have been omitted. The numbering of the amino acid sequence corresponds to that used in the text although it should be noted that there are uncertainties in the exact number of residues at the N- and C-termini and at the "gaps" in the electron density between residues 51-57, 164-169 and 205-208, marked in the figure by broken connections. (After Fenna et al., 1977.)

X-ray sequence and describe those aspects of the structure that do not depend critically on the correct identification of individual amino acids, i.e. the backbone conformation and the configuration of the bacteriochlorophylls.

Methods

Data Collection

Crystals of the bacteriochlorophyll *a* protein have space group $P6_3$ with cell dimensions *a* = *b* = 111.9Å, *c* = 98.6Å (Olson *et al.*, 1969; Matthews *et al.*, 1979). In the initial X-ray analysis, precession photography was used to obtain an electron density map at 2.8Å resolution. This resolution is essentially the highest that could be achieved using conventional precession photographs, a limitation that can be attributed in part to the relatively large volume of the unit cell ($1.08 \times 10^6 Å^3$) and in part to the high (70%) solvent content of the crystals. By using oscillation photography and monochromatic radiation (Schmid *et al.*, 1981) it has been possible to extend the resolution of the data to 1.9Å. A total of 9 crystals was used to obtain 76 oscillation photographs. These were processed using a program based on that of Rossmann (1979). A total of 126,000 reflections were measured and reduced to a unique set of 46,700 intensities. The average agreement between symmetry-related intensities on the same film was 4.6% and the overall agreement between reflections measured on different films was 11.6%. (The latter value is much higher than the former in part because the reflections come from different crystals in different

orientations, but also because the number of measurements per reflection is larger.) For the 15,000 reflections common to the precession and oscillation data sets the agreement was 9.8%.

Starting Coordinates

A model of the structure was built in an optical comparator (Richards, 1968; Colman et al., 1972). Raw coordinates were then obtained by placing markers in the electron density map to correspond with the model and the density. These coordinates were then adjusted to have acceptable stereochemistry and were the starting point for the structure refinement. The initial crystallographic R-value was 44.2% for data between 6.0Å and 1.9Å resolution.

Refinement

The protein was refined using a new, fast, flexible, restrained least-squares refinement package developed in collaboration with L.F. Ten Eyck (see Chapter III). The program restrains bond lengths and bond angles to be within acceptable limits. Also, groups of atoms known to be planar (e.g. aromatic side chains, carboxyl, peptide and amide groups) are likewise restrained. Distances between non-bonded atoms within the protein, between neighboring molecules in the crystal and between protein and solvent atoms are routinely checked and can be restrained if desired.

Although it was expected that the 22 conjugated atoms in each of the dihydrophorbin macrocycles would be essentially coplanar, it was

still necessary to allow for the possibility that the Bchl head-groups might be distorted in one way or another. Therefore instead of constraining all the conjugated atoms to be near a single plane, we divided up the atoms into sets of smaller subplanes that had common overlapping atoms but were otherwise independent. This procedure ensured that planarity was maintained locally but allowed larger-scale deviations. The subgroups of atoms were as follows: (C1C, CHC, C4B, C3B, C2B, C1B, CHB, C4A, NA, C1A, CHA, C4D), (C1A, CHA, C4D, CBD, C2D, C1D, CHD, C4C, NC, C1C, CHC, C4B), (NB, C1B, C2B, C3B, C4B), (ND, C1D, C2D, C3D, C4D) and (CAD, OBD, CBD, C3D).

Because the bacteriochlorophyll _a_ protein was used as a test case for development of the refinement package the refinement proceeded in a somewhat irregular manner. Also, in the early stages a graphics system was not available and the model was checked by direct inspection of plotted contour maps. By these methods the crystallographic residual was reduced to about 23%. In the latter stages an MMS-X graphics system (Molnar _et al._, 1976) running the program FRODO (Jones, 1982) was used to facilitate the inspection of both ($2F_o - F_c$) and ($F_o - F_c$) maps. Also ($2F_o - F_c$) "omit" maps were calculated in which 10 amino acids at a time were omitted from the phase calculation and their density checked for consistency with the model. (Our experience was that such "omit" maps are not obviously superior to conventional difference maps.) We adopted the conservative policy of deleting atoms from the model rather than having them present in doubtful positions. Also we only allowed side chains that corresponded to the standard amino acids. Thermal

factors were refined in cycles independent of coordinate refinement. The presence of atoms with large thermal factors was taken as an indication that the identification of the amino acid in question should be changed. Solvent atoms with thermal factors greater than 50Å^2 were deleted.

As additional checks, the refinement program provides a list of the worst occurrences of each type of stereochemistry including bond lengths, bond angles, torsion angles, inverted chiral centers and, if desired, a list of the atoms with the largest crystallographic gradients. These lists were routinely used to check for possible errors in the model.

Altogether, five rounds of manual model revision and automated refinement were completed before convergence was achieved. The final R-value is 18.9% to 1.9Å resolution. In the final model the bond lengths and angles depart from their ideal values by root mean square values of 0.02Å and 3.2° respectively. The overall shifts in the polypeptide backbone during refinement are shown in Figure 5. The refined coordinates have been deposited in the Brookhaven Data Bank.

Results and Conclusions

Amino Acid Sequence

There are several difficulties in attempting to determine an amino acid sequence from crystallographic data. In other instances it has proved possible to correctly identify about 50-75% of the amino acids by X-ray methods (e.g. see Matthews, 1977). Several pairs of amino acids

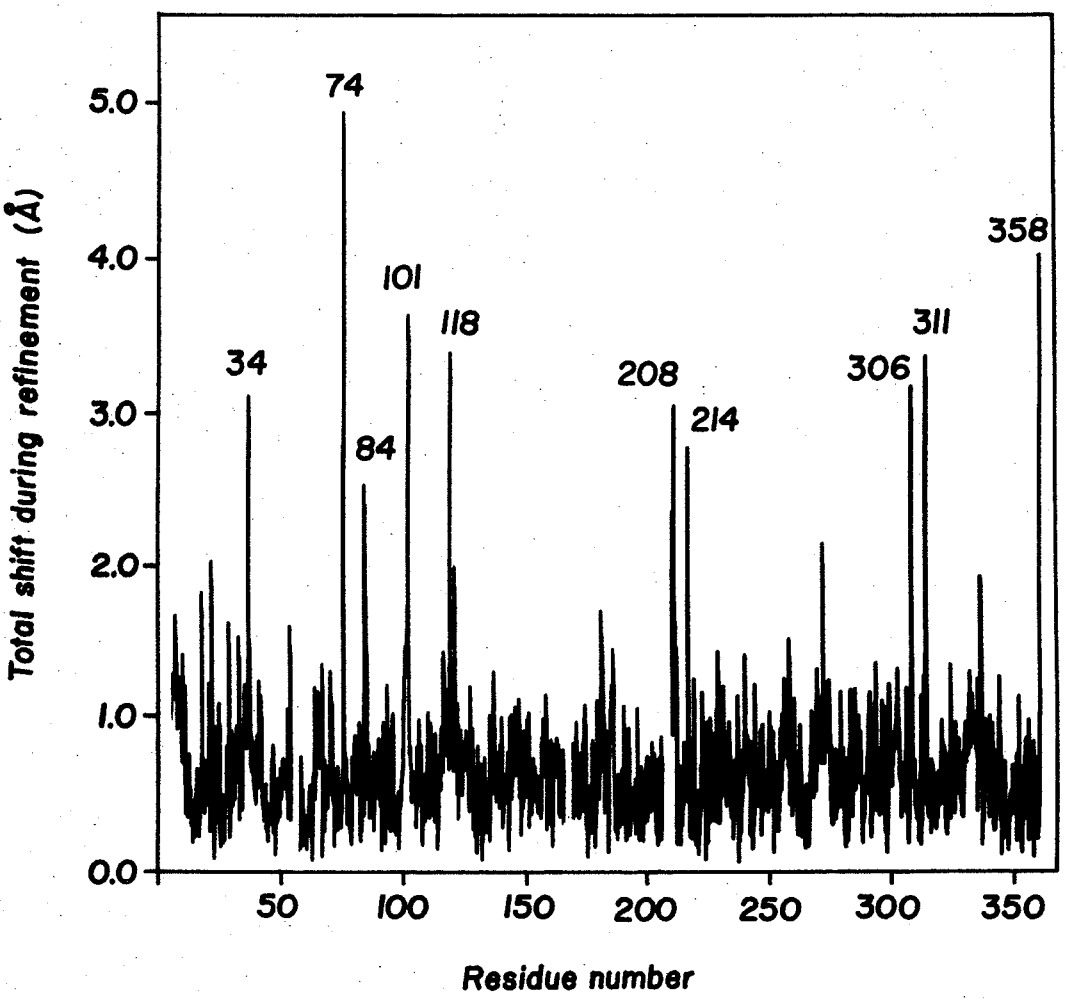


FIGURE 5. Coordinate Shifts of the Backbone Atoms of the Bchl-Protein During Refinement

The shows, for each backbone atom (N, CA, C, O), the difference between the position at the start of refinement and in the final model. Atoms that have undergone the largest shifts are numbered by their residue number.

are chemically different but isostructural, namely (Val, Thr), (Glu, Gln) and (Asp, Asn). Members of these pairs are virtually impossible to differentiate in the electron density map. In addition, some amino acid side chains, especially on the surface of the protein, are mobile and hard to identify. In extreme cases the backbone atoms also undergo large "thermal" displacements and cannot be seen. Therefore we have given a confidence rating, ranging from 1 to 4, for each amino acid identification (Table 3). Very badly determined residues were given a rating of "4". An amino acid in this category did not fit the existing electron density well, had questionable stereochemistry and its atoms usually had high B values. Thirty-six amino acids were determined to belong in this category. If a residue did not agree with the electron density but the disagreement was not so bad as to eliminate the possibility of the identity being correct it was rated as "3". Forty-eight amino acids fell into this class. If an amino acid had some disagreement with the electron density but was still considered likely to be correct it was rated "2". Forty-two amino acids fit this description. The final class (rating "1") was reserved for amino acids that agreed well with all the criteria. Two hundred eighteen amino acids were so classified.

The sequence confidence rating scheme is unquestionably subjective. In the case of the isostructural pair Thr and Val, the confidence rating does not take into account the ambiguity between these residues. I.e. if a residue appeared to be a good Val or Thr it was considered to be well determined even though the two alternatives could not be reliably

TABLE 3. "X-ray" Amino Acid Sequence[a]

1									10										20
---	---	Val 2	Ser 3	Ala 1	Asx 4	Ser 1	Ala 4	Tyr 1	Lys 4	Ile 1	Ile 1	Leu 1	Glx 3	Gly 1	Gly 2	Ala 4	Ser 1	Ser 1	Trp 1
21									30										40
Gly 1	Glx 3	Val 1	Ala 3	Gly 1	Ala 3	Ala 2	Ser 3	Val 1	Ser 4	Val 1	Pro 1	Ala 1	Ser 3	Ile 1	Pro 1	Leu 1	Asx 3	Pro 1	Thr 1
41									50										60
Asx 3	Cys 1	Ser 3	Ile 1	Lys 2	Ile 1	Asx 1	Ala 1	Ser 4	Pro 1	Ser 4	---	---	---	---	---	Thr 4	Val 1	Lys 1	Phe 1
61									70										80
Thr 4	Val 1	Ala 3	Ile 1	Ala 4	Ser 1	Thr 1	Ile 3	Asx 1	Ala 3	Thr 1	Ala 2	Asx 1	Thr 2	Leu 3	Ser 4	Val 1	Ala 4	Thr 2	Ser 3
81									90										100
Ile 1	Ala 1		Asx 1	Ser 4	Ala 3	Ala 3	Ser 3	Lys 1	Arg 1	Ile 1	Ala 1	Thr 1	Gly 1	Ala 4	Gly 1	Ser 2	Val 4	Ala 3	Val 1
101									110										120
Gly 1	Ser 4	Phe 1	Ala 3	His 1	Ala 1	Phe 2	Ser 1	Phe 2	Met 3	Gly 1	Ser 2	Thr 1	Thr 1	Asx 1	Met 4	Tyr 1	Tyr 1	Ser 4	Gly 4
121									130										140
Ser 4	Ser 4	Ala 1	Thr 1	Ala 2	Arg 1	Asx 1	Ile 1	Pro 1	Asx 1	Pro 1	Ile 1	Tyr 1	Met 3	Glx 1	Gly 1	Arg 1	Glx 1	Phe 1	His 1
141									150										160
Asx 1	Ile 1	Asx 1	Met 1	Lys 1	Val 1	Pro 1	Leu 1	Asx 3	Asx 2	Gly 3	Asx 2	Leu 1	Thr 1	Ser 4	Thr 1	Trp 1	Lys 3	Gly 1	Phe 2

TABLE 3. (continued)

161									170									180
Ser	Ala	Ala	Ser	---	---	---	---	Asx	Phe	Gly	Asx	Trp	Ile	Arg	Asx	Phe	Trp	Phe	Ile
4	4	1	4					1	1	1	1	1	1	1	3	1	1	1	1
181									190									200
Gly	Pro	Ala	Phe	Ala	Ala	Ile	Asx	Glx	Gly	Gly	Glx	Arg	Ile	Ser	Pro	Val	Thr	Thr	Asx
1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	3	1	2	1
201									210									220
Ser	Ala	Ser	Thr	Glx	---	---	Gly	Pro	Asx	Gly	Thr	Thr	Arg	Trp	Ser	Phe	Ser	His	Ala
2	2	3	2	4			4	3	4	1	1	1	1	1	4	1	2	1	1
221									230									240
Gly	Ser	Gly	Val	Val	Asx	Ser	Ile	Ser	Arg	Trp	Thr	Glx	Leu	Phe	Pro	Thr	Ala	Lys	Leu
1	1	1	2	1	1	4	1	2	1	1	1	1	1	1	1	2	4	1	1
241									250									260
Ser	Lys	Pro	Ala	Ala	Ile	Glx	Gly	Gly	Phe	Ser	Ser	Asx	Ser	Ala	Gly	Ile	Ser	Val	Ala
3	1	1	1	2	1	4	1	1	1	3	1	2	2	2	1	1	4	1	1
261									270									280
Val	Ala	Gly	Ser	Leu	Pro	Gly	Val	Ser	Lys	Ser	Ala	Gly	Gly	Gly	Asx	Lys	Lys	Ile	Leu
1	4	1	3	1	1	1	1	1	2	3	2	3	3	3	2	2	1	1	1
281									290									300
Asx	His	Pro	Asx	Ile	Pro	Leu	Thr	His	His	Gly	Met	Thr	Gly	Lys	Phe	Asx	Ser	Phe	Ser
1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	3
301									310									320
Ser	Asx	Thr	Ala	Asx	Lys	Ile	Thr	Leu	Pro	Lys	Gly	Tyr	Ala	Ile	Ser	Tyr	Ala	Ala	Pro
3	1	1	3	2	1	1	1	1	1	3	1	1	2	2	3	1	1	1	1

TABLE 3. (continued)

321									330									340
Ala	His	Ser	Ser	Lys	Asx	Asx	Glx	Ala	Tyr	Lys	Trp	Ala	Gly	Gly	Ala	Tyr	Ala	Arg	Trp
2	2	3	2	3	2	1	2	4	1	1	1	2	1	2	3	1	1	1	1

341									350								358		
Val	Glx	His	Val	Cys	Lys	Gly	Gly	Thr	Gly	Glx	His	Glx	Thr	Leu	Tyr	Ala	Ala		
1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	3	4		

(a) The amino acid sequence is inferred from the crystallographic refinement. It is not possible to distinguish Thr and Val directly. The identifications given here are inferred indirectly from the environment. Numbers under each amino acid indicate the level of confidence in the identification (1, most confident; 4 least confident). The exact number of amino acids at the N- and C-termini and in the "gaps" is uncertain.

distinguished. In such cases an attempt was made to differentiate between the two alternatives by consideration of the local environment, but this is, at best, an indirect assessment. As far as possible we kept the numbering of the amino acid sequence in register with that inferred from the original isomorphous replacement electron density map. Therefore, no particular significance should be placed on the exact numbering of residues in the "gaps" or at the amino terminus. For example, the X-ray sequence has a "gap" between residues 51 and 57 but this is not meant to imply that there are thought to be exactly 5 intervening residues. Also there is no residue corresponding to sequence number 83. Compared to the preliminary X-ray amino acid sequence obtained from inspection of the isomorphous replacement electron density map (Matthews et al., 1979), 179 amino acids remain the same and 165 were altered during refinement.

Amino Acid Composition

The amino acid composition of the Bchl a protein is known (Olson et al., 1976) and is compared with the amino acid composition of the X-ray model in Table 4. The X-ray model contains 344 residues and is presumed to be missing several additional residues (e.g. at "gaps" in the electron density and possibly at the ends of the polypeptide chain). Therefore the amino acid composition determined chemically and given in Table 4 has been put on a scale such that it corresponds to a total composition of 360 residues.

Table 4 indicates that there are too few glycines in the X-ray model. This is somewhat surprising since one tends to put glycines in

TABLE 4. Amino Acid Composition

Amino Acid	Experimental[a]	X-ray Model	Discrepancy
Gly	41	36	-5
Asx	40	31	-9
Val	34-35	18	-16-17
Glx	34	13	-21
Ser	28	47	19
Ile	24	23	-1
Ala	22	48	26
Leu	20	12	-8
Arg	20	8	-12
Pro	18	17	-1
Lys	18	17	-1
Phe	17	15	-2
Thr	14	26	12
Tyr	10	9	-1
His	8	9	1
Trp	7	8	1
Met	3-4	5	1-2
Cys	2	2	0
Total	360	344	-16

[a] From Olson *et al.* (1976), scaled so that there are a total of 360 amino acids per polypeptide.

regions of the model that are poorly determined. It is possible that there are additional glycines, not included in the model, that are in flexible bend or terminal regions. It is not surprising that the model has too many serines and alanines since these could be surface residues for which the distal part of the side chain is mobile and not seen in the electron density map. In particular, the 45 "extra" serines and alanines in the model could well correspond to many of the 42 "missing" Asx's, Glx's and Arg's which would be expected to be on the surface of the molecule.

Backbone Conformation

The refinement supports the overall conformation of the protein backbone reported previously (e.g. see Figs. 5, 6, and 7 of Matthews et al., 1979). Figure 6 shows the Ramachandran diagram for the backbone dihedral angles. These angles were not monitored or manually adjusted during the refinement to conform to "acceptable" limits. There are four non-glycine residues that occur in regions where residues with beta-carbons are not normally expected. Residues 69 and 70 are the central two residues in a Type I' turn (Venkatachalam, 1968). Amino acid 69 has been identified with reasonable confidence as Asx and while the identity of residue 70 is not clear it seems to be at least as large as alanine. The turn is distorted with the carbonyl oxygen of residue 68 accepting the bridging "turn" hydrogen bond as well as a second hydrogen bond from a water molecule as is often associated with these hairpin bends (Rose et al., 1983). Residue 327 is also an apparent

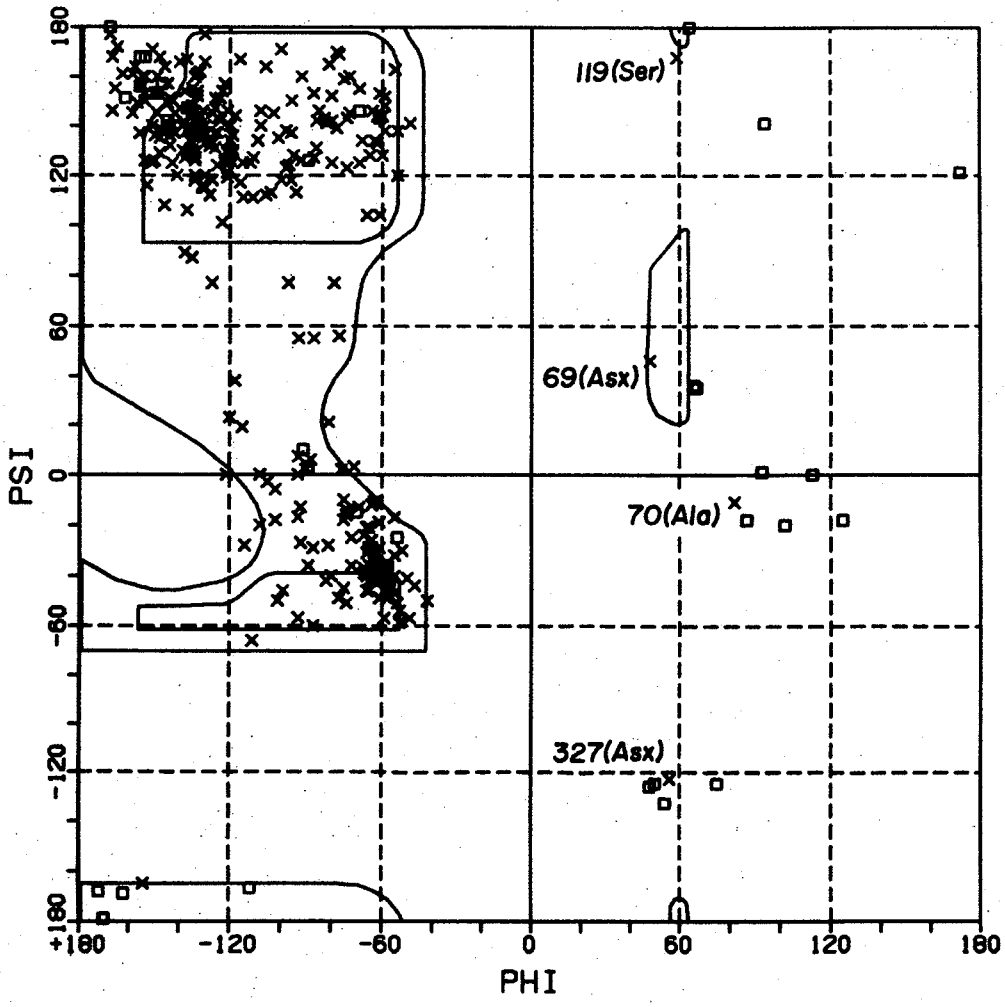


FIGURE 6. Ramachandran Diagram for the Bchl-Protein

Presumed glycine residues indicated by square symbols and non-glycine residues by crosses.

non-glycine residue (Asx) in a turn (Type II'). Residue 119 is within a poorly defined stretch of chain and the conformational angles for residues in this vicinity are unreliable.

In the present numbering scheme for the amino acid sequence, residues that form α-helices are 121-128, 150-164, 169-179, 183-190, 225-230, 230-234, 284-294 and 335-346. One of the characteristics of the Bchl protein is the large number of residues in extended conformations. These residues form two β-sheets, one of 16 strands and the other of 4 strands (Matthews et al., 1979). The β-sheet strands include residues 3-14, 20-29, 39-49, 57-68, 71-85, 88-101, 103-118, 135-146, 191-205, 209-222, 244-253, 256-265, 268-272, 276-280, 299-309, 314-319, 324-326, 328-333, and 352-356. (Residues are included in α-helices and β-sheets if they contribute one or more hydrogen bonds to the secondary structure element.) There is an extended six-residue segment (278-283) that begins within one of the β-sheet strands and continues beyond in a conformation similar to that of collagen (Soman & Ramakrishnan, 1983).

In four instances the main chain forms "1-3" hydrogen bonds between the carbonyl oxygen of one peptide and the nitrogen of the succeeding peptide (e.g. see Ramachandran & Sasisekharan, 1968; Baker & Hubbard, 1984). Two examples are shown in Figures 7(a) and 7(b). In all cases the Ramachandran angles for the α-carbon enclosed within the H-bonded peptides are in the vicinity of $\phi \sim -90°$, $\psi \sim 70°$ (Table 5).

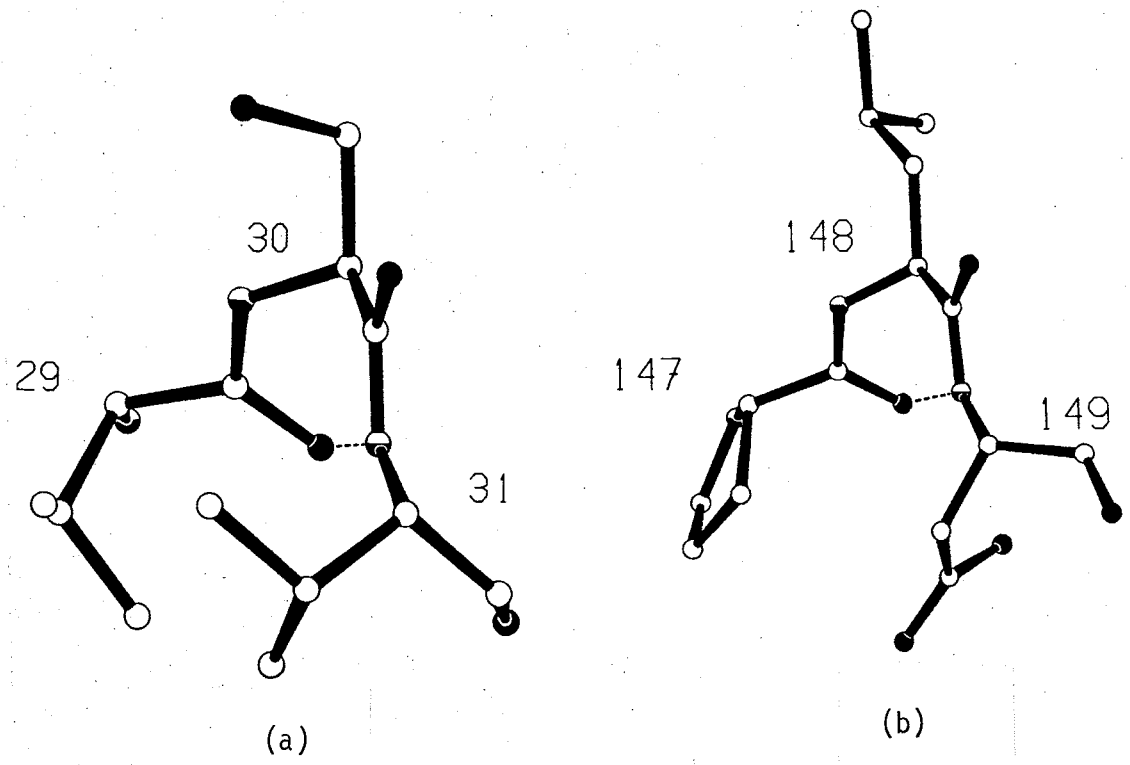


FIGURE 7. Two Examples of "1-3" Hydrogen Bonds Between Successive Peptide Groups

(a) The presumed sequence Val 29 - Ser 30 -Val 31. (b) The presumed sequence Pro 147 - Leu 148 - Asx 149.

TABLE 5. "1-3" Hydrogen Bonds

Residue	ϕ(°)	ψ(°)	d(Å)	Angle (°) C-O•••N
30 (Ser)	-88	58	3.1	84
148 (Leu)	-94	55	3.1	82
179 (Phe)	-98	80	3.2	88
240 (Leu)	-80	76	2.8	98

Dynamics

The apparent motion of the polypeptide backbone as estimated from the refined crystallographic thermal factors is shown in Figure 8. The three places where there are "gaps" in the backbone (i.e. residues 52-56, 165-168, and 206-207) are, as expected, regions of large apparent thermal motion or disorder. Often the regions of high mobility are at the ends of "hairpin loops" between successive β-strands, but there is also a case (near residue 164) where one end of an α-helix is very mobile. This α-helix is at the left edge of the subunit shown in Figure 4. It lies against the surface of the protein but is largely exposed to solvent.

The residues of the protein that form ligands to the magnesium atoms of the bacteriochlorophylls tend to have relatively low mobility (Fig. 8). This is presumably due to the fact that such residues are internal, and should not be taken to imply that the interactions with the metals "anchor" the protein.

The thermal factors for both the bacteriochlorophyll head groups and the phytyl chains are relatively low, showing that these atoms are as well ordered as the interior protein atoms. In particular, the refinement confirms that the phytyl chains have well-defined conformations (Matthews et al., 1979; Matthews, 1982) and are not disordered as in a fluid lipid bilayer.

Bacteriochlorophyll a Conformations

The atom nomenclature used here for the bacteriochlorophyll a molecules is illustrated in Figure 9. This scheme follows the

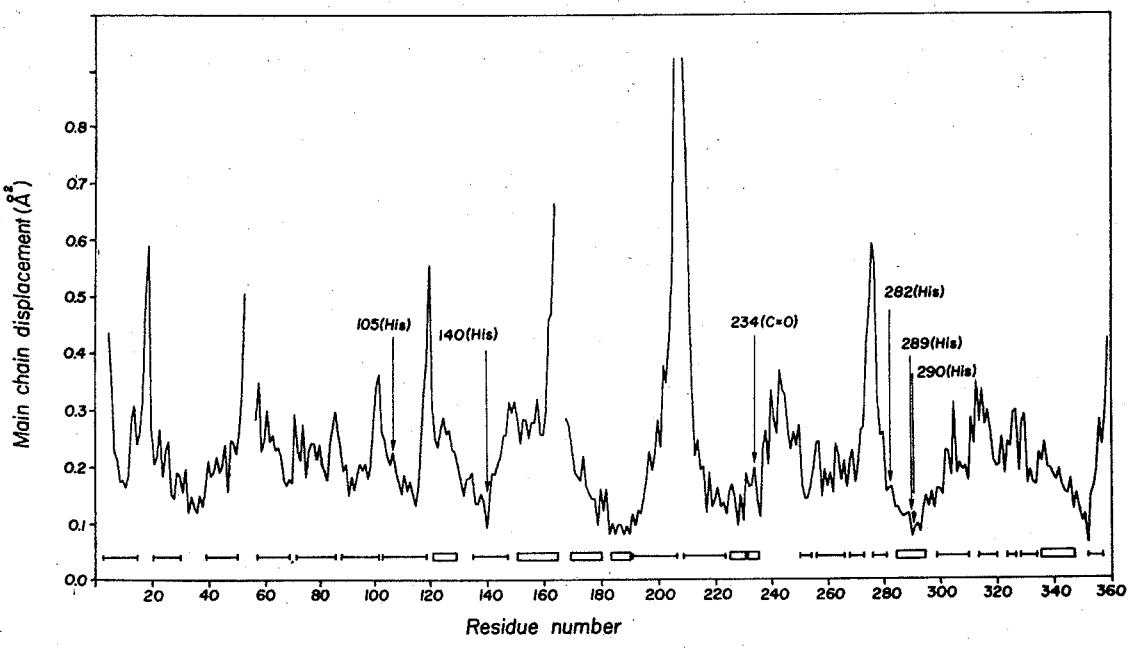


FIGURE 8. Apparent Thermal Motion of the Backbone of the Bchl-Protein

There are three breaks in the chain at residues 52-56, 165-168 and 206-207 where the electron density is too weak to follow. Locations of α-helices are indicated by open rectangles and β-sheet strands by single bars. The residues that are liganded to the bacteriochlorophyll magnesiums are indicated.

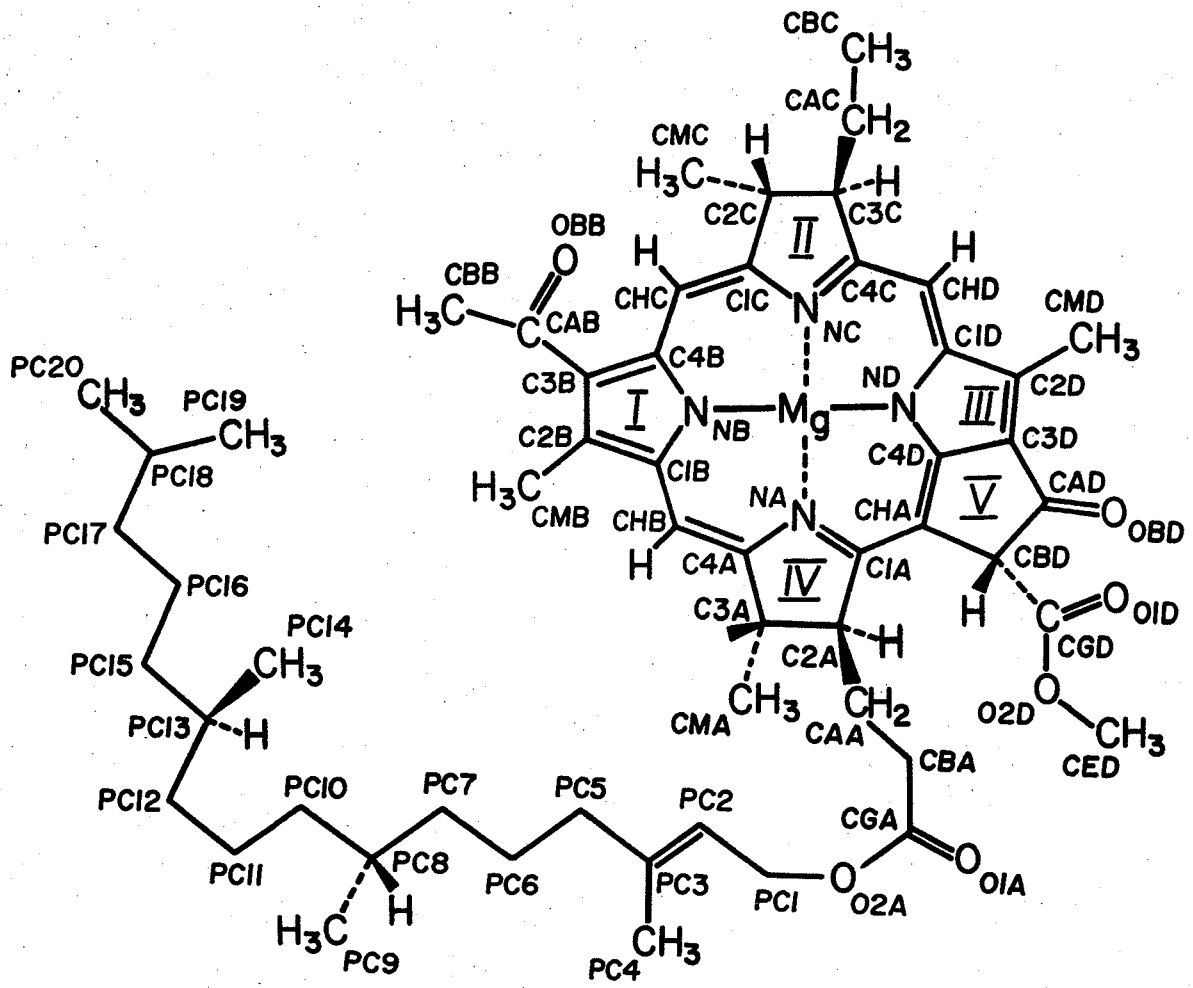


FIGURE 9. Schematic Drawing of the Structure of Bacteriochlorophyll a Showing the Scheme Used for Atom Identification

It should be noted that the scheme is somewhat anomalous in that atoms associated with Rings I, II, III, IV or V have the letter B, C, D, A or D, respectively as the last character in the atom name. We have changed to this scheme because it is the one used by the Brookhaven Data Bank.

conventions used by the Brookhaven Data Bank, and replaces the nomenclature used previously (Fenna et al., 1977; cf. Strouse, 1974).

There are seven bacteriochlorophyll a molecules in the Bchl protein. This is a reasonable sample within which to compare variables such as the out-of-plane distances of the magnesiums, the pucker of rings II and IV, possible deformations of the chlorin rings and variability in the conformations of the ring substituents.

In Figure 10 the porphine head-groups of all seven Bchls have been superimposed by a least squares process (Kabsch, 1976). The figure gives an overall impression of the variability in the respective conformations. One of the striking features is the distribution of the magnesium positions. In two cases (Bchls 3 and 7) the metal ions are "below" the ring plane whereas in the other five cases the magnesiums are "above". There are no intermediate states.

Table 6 gives, for each Bchl, the distances of all the magnesiums from the planes with best least-squares fit to the 22 conjugated atoms. The table indicates that the magnesiums of Bchls 3 and 7 have a mean displacement of 0.48Å below the plane of the ring whereas the other five magnesiums are, on average, 0.48Å above. These distances agree reasonably well with the value of 0.4Å observed in the structure of ethylchlorophyllide a (Strouse, 1974; Chow et al., 1975; Kratky & Dunitz, 1975).

The displacement of each of the magnesiums is always toward the side of the ring plane from which the metal is liganded to the protein. (Each magnesium is pentacoordinate.) Therefore the protein environment

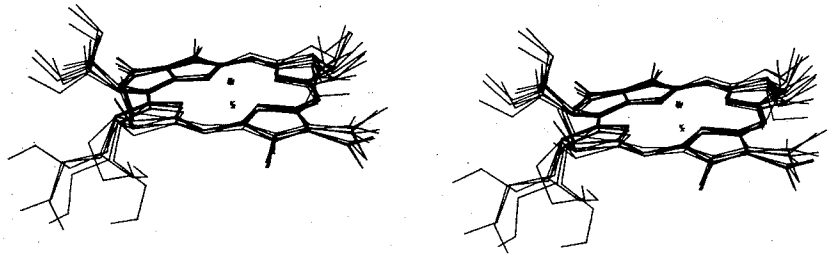


FIGURE 10. Stereo Drawing Showing the Head-Groups of All Seven Bchls Brought Into the Same Orientation

The "X's" indicate the positions of the magnesiums.

TABLE 6. Magnesium:Ligand Distances and Magnesium Out-of-Plane Distances

Bchl	Presumed Mg Ligand	Mg-ligand Distance (Å)	Mg out-of-plane Distance (Å)
1	His 105:NE2	2.2	0.44
2	Solvent (SOL1:14S)	2.0	0.45
3	His 290:NE2	2.1	-0.54
4	His 282:NE2	2.1	0.46
5	Leu 234:O (Backbone)	2.0	0.52
6	His 140:NE2	2.2	0.54
7	His 289:ND1	2.1	-0.43

determines the direction of the out-of-plane magnesium displacement; the displacement direction is not an intrinsic property of the bacteriochlorophyll molecule itself.

The protein ligands for the two Bchls with their magnesiums below the ring plane appear to be histidines although one of these (Bchl 7) is unique in that it is the ND1 imidazole nitrogen that binds to the metal. For the other four Bchls with presumed histidine ligands the magnesium is liganded by nitrogen NE2. Table 6 lists the presumed ligands for each of the Bchls together with the ligand distances.

For the Ring I acetyl substituent, the oxygen and the methyl group are essentially equivalent in the X-ray refinement. Therefore these moieties have to be differentiated indirectly from their presumed interactions with the protein. These are summarized in Table 7. During refinement, the acetyl group was constrained to be planar but was allowed free rotation about the bond between C3B and CAB. Based on a resonance Raman study, Lutz et al. (1982) suggested that at least four and possibly all seven of the acetyl groups should be within 18° of coplanar with the conjugated ring system. Allowing for experimental error, this is as observed (Table 7; Fig. 11(a)). Five of the acetyl groups are within 24° of coplanarity and the greatest departure (Bchl 5) is 36°. The torsion angles (Table 7) fall into two classes differing by approximately 180°. In two cases the ester oxygen does not make any apparent hydrogen bond to the protein and so cannot be distinguished from the methyl group. In such cases the torsion angle could be χ or $\chi - 180°$. It is worth noting that Lutz et al. (1982) also proposed from

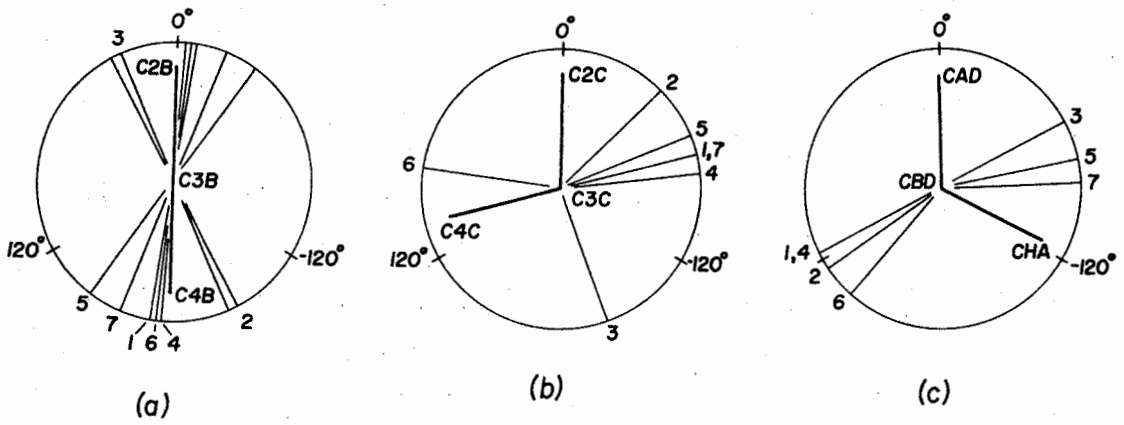


FIGURE 11. Newman Projections Illustrating the Torsion Angles of the Ring Substituents for the 7 Bchls

In each case the direction of view is from the substituent toward the Bchl ring. (a) Ring I acetyl groups. The direction of view is along the bond CAB-C3B and the lines labeled 1-7 indicate the planes of the acetyl groups of the 7 Bchls. In this figure no attempt is made to distinguish between the acetyl oxygen (OBB) and the acetyl methyl group (CBB) (cf. Table 7). (b) Ring II ethyl groups. The direction of view is along the bond CAC-C3C and the numbers 1-7 indicate, for the 7 Bchls, the torsion angles about this bond (cf. Table 8). (c) Ring V carbomethoxy groups. The view direction is along the CGD-CBD bond and the numbers 1-7 indicate the apparent torsion angles (defined by atom O2D) for the 7 Bchls (cf. Table 9).

TABLE 7. Configurations of the Ring I Acetyl Groups

Bacteriochlorophyll	Dihedral Angle[a]	Atom H-bonded to OBB[b]	H-bonded Distance (A)
1	171°	Solvent (SOL3:410)	2.6
2	-151°	Ser 66:OG	2.8
3	24°	Tyr 9:OH	2.7
4	176° or -4°	--	--
5	144° or -36°	--	--
6	174°	Trp 176:NE1	3.0
7	158°	Solvent (SOL1:43S)	2.7

(a) The dihedral angle is the angle between the planes defined by (C2B,C3B,CAB) and (C3B,CAB,OBB) and is 0° when C2B and OBB are cis. The calculation assumes that atoms OBB and CBB can be differentiated (see text).

(b) The identification of these atoms is based only on the X-ray results and should be regarded as very tentative pending confirmation with the chemically determined amino acid sequence.

their Raman study that up to two of the acetyl carbonyls were free of hydrogen bonding, in agreement with Table 7. The acetyl group with greatest departure from coplanarity with the chlorin system is not hydrogen bonded but has its methyl group "wedged" between hydrophobic groups of the protein. These close contacts appear to prevent the acetyl group relaxing to the coplanar configuration.

The torsion angles for the Ring II ethyl substituents are listed in Table 8 and shown in Figure 11(b). The conformations are invariably staggered with one value in the 60° region, a second in the 180° region but a strong preference for torsion angles near -60° (five cases out of seven).

In refining the Ring V carbomethoxy groups, atoms CBD, CGD, OD1 and OD2 were constrained to lie in a common plane but free rotation was allowed about the CBD-CGD bond. Even though the methyl group CED was not restrained to be coplanar with the remainder of the carbomethoxy moiety, it does lie close to the common plane in six of the seven cases (Fig. 10). The exception is Bchl 6 where the methyl group appears to be held away from coplanar by steric interference with neighboring protein side chains. The plane of the carbomethoxy group is, in all cases, essentially perpendicular to the plane of the conjugated chlorin ring system (Fig. 11(c); Table 9). In no case does the ester oxygen O2D appear to accept a hydrogen bond from the protein. (This is also the case for the ester oxygen O2A in each of the phytyl chains.)

In each of the seven Bchls the keto oxygen (OBD) is essentially coplanar with the remainder of the conjugated system (Table 11). Five

TABLE 8. Configurations of the Ring II Ethyl Groups

Bacteriochlorophyll	Torsion Angle[a]
1	-75°
2	-45°
3	-159°
4	-83°
5	-67°
6	82°
7	-75°

[a] The torsion angle is zero when C2C and CBC are cis.

TABLE 9. Configurations of the Ring V Carbomethoxy Groups

Bacteriochlorophyll	Dihedral Angle[a]	Atom H-bonded to O1D[b]	H-bond Distance (Å)
1	118°	--	--
2	124°	--	--
3	-62°	--	--
4	118°	Solvent (SOL1:26S)	2.7
5	-78°	--	--
6	139°	Thr 113:OG1	3.1
7	-88°	--	--

[a] The dihedral angle is 0° when CAD and O2D are cis.

[b] Identifications provisional pending confirmation with the chemically determined amino acid sequence.

of the keto oxygens appear to accept hydrogen bonds from the protein or a bound solvent molecule, while two seem not to be hydrogen bonded (Table 10). The resonance Raman spectra suggest that up to three of the keto oxygens are free from hydrogen bonding (Lutz et al., 1982).

Each Bchl head group is connected to its phytyl tail via a propionic acid linkage. The conformations of the phytyl chains are all different and the linking groups also have varied conformations. As mentioned, none of the linking oxygens O2A accept hydrogen bonds. Also in only one case (Bchl 1) does oxygen O1A appear to make a hydrogen bond and this is to an atom in one of the symmetry-related monomers of the Bchl protein trimer.

There are a series of close contacts between the oxygen-containing ring substituents of Bchls 2 and 7. In the carbomethoxy group of Bchl 7, oxygen O1D is 3.0Å from oxygen O2A in the same Bchl. At the same time the other oxygen (O2D) in the carboxymethyl group of Bchl 7 is 3.0Å from oxygen O2A in Bchl 2. None of these close approaches are hydrogen bonds since no protons are involved.

Although the 22 atoms that constitute the conjugated part of each of the Bchl head-groups are essentially coplanar (Table 11) it can be asked if there is any evidence for systematic distortions away from planarity. Such an analysis is complicated because the relatively large uncertainty in the coordinates of the conjugated atoms (ca 0.15Å) could well mask small but systematic departures from planarity. Our first step, therefore, was to see if there were systematic departures from planarity that were common to two or more of the Bchls.

TABLE 10. Interactions of the Ring V Keto Oxygens

Bacteriochlorophyll	Interacting Atom[a]	Distance (Å)
1	--	--
2	Peptide NH:Ile 132[b]	2.9
	Peptide NH:Tyr 133[b]	3.1
3	Peptide NH:Ser 34	2.7
	OG:Ser 34	2.9
4	OH:Tyr 337	2.9
5	Solvent (SOL1:41S)	3.0
6	NH2:Arg 90	3.0
7	--	--

(a) All identifications tentative pending confirmation with the chemically determined amino acid sequence.

(b) Symmetry-related molecule.

TABLE 11. Departures from Planarity of the Seven Bchl Head-Groups

The entries give the distance in Angstroms from the best least-squares plane through the 22 atoms (asterisked) constituting the conjugated part of the dihydrophorbin macrocycle.

	Bchl 1	Bchl 2	Bchl 3	Bchl 4	Bchl 5	Bchl 6	Bchl 7
MG	0.44	0.45	-0.54	0.46	0.52	0.54	-0.43
C1A*	-0.05	-0.24	-0.19	0.05	0.03	0.05	-0.05
C2A	-0.01	-0.27	-0.20	0.03	0.08	-0.31	-0.01
CAA	-1.09	-1.71	-1.57	-1.34	-1.35	-1.77	-1.42
CBA	-2.59	-1.80	-1.20	-2.31	-1.17	-2.69	-1.57
CGA	-3.00	-3.21	-2.51	-3.01	-2.48	-4.12	-0.71
O1A	-3.05	-3.88	-3.18	-3.04	-2.54	-4.39	-0.94
C3A	-0.14	0.08	0.24	0.62	0.44	0.14	0.39
CMA	0.94	1.42	1.72	2.09	1.88	1.51	1.84
C4A*	-0.02	0.01	-0.02	0.16	0.27	0.24	0.17
NA*	0.00	-0.18	-0.18	0.11	0.01	0.19	0.04
CHB*	-0.06	0.09	0.13	0.26	0.01	0.02	0.07
C1B*	0.10	0.22	0.15	0.04	-0.06	-0.02	0.07
C2B*	0.06	0.11	0.18	-0.10	-0.11	-0.24	-0.21
CMB	-0.26	-0.05	0.23	-0.33	-0.03	-0.29	-0.35
C3B*	-0.02	-0.03	0.13	-0.24	-0.07	-0.28	-0.20
CAB	-0.04	0.04	0.21	-0.56	0.25	-0.65	-0.46
OBB	-0.28	0.42	-0.08	-0.71	-0.28	-0.61	-0.80
CBB	0.36	-0.23	0.54	-0.70	1.14	-0.77	-0.32
C4B*	0.00	-0.03	0.00	-0.20	-0.02	-0.05	0.05
NB*	0.04	0.17	0.03	0.01	0.01	0.14	0.22
CHC*	0.02	-0.13	-0.24	-0.14	-0.17	-0.14	-0.07
C1C*	0.04	-0.07	-0.14	-0.03	0.04	0.05	0.07
C2C	0.16	-0.33	-0.32	-0.13	0.24	0.01	0.51
CMC	1.59	0.70	0.46	0.78	1.61	1.10	1.96
C3C	-0.24	-0.09	0.30	0.24	0.06	0.15	0.63
CAC	-1.72	-1.15	-0.51	-0.85	-1.28	-1.03	-0.14

TABLE 11. (continued)

CBC	-1.97	-1.45	0.17	-0.45	-1.24	-2.37	0.46
C4C*	-0.09	-0.09	0.03	0.21	0.06	0.08	0.10
NC*	0.02	0.00	-0.05	0.08	0.07	0.18	0.01
CHD*	-0.06	0.12	0.17	0.29	0.03	0.08	0.02
C1D*	0.07	0.04	-0.02	-0.03	0.04	-0.01	-0.07
C2D*	0.00	0.04	0.12	-0.07	0.00	-0.11	0.01
CMD	-0.07	0.35	0.30	-0.05	0.04	-0.16	-0.05
C3D*	-0.08	0.03	0.07	-0.03	0.03	-0.02	-0.01
CAD*	0.04	-0.02	-0.05	-0.19	-0.14	-0.25	-0.08
OBD*	0.20	0.22	0.24	-0.18	-0.16	-0.21	0.03
CBD*	-0.25	-0.33	-0.28	-0.21	-0.06	-0.32	0.00
CGD	0.87	0.65	0.75	1.04	1.09	0.79	1.18
O1D	2.04	1.82	0.60	2.17	0.93	1.90	1.04
O2D	0.49	0.06	1.97	0.76	2.30	0.55	2.44
CED	1.67	0.99	2.96	1.97	3.48	1.89	3.51
CHA	-0.13	-0.18	-0.19	0.01	-0.08	0.04	-0.06
C4D*	-0.08	-0.03	-0.07	-0.01	0.08	0.12	-0.02
ND*	0.01	-0.03	-0.11	0.00	0.11	0.15	-0.09

Table 12 shows the correlation coefficients between the out-of-plane displacements of the 22 conjugated atoms calculated for all possible pairs of Bchls. The striking result is that some pairs of Bchls (e.g. 2 and 3; 4 and 6; 5 and 6) have correlation coefficients close to unity, indicating that they depart from planar in virtually the same way. There are some instances of negative correlation (e.g. 1 and 4; 3 and 6) indicating that these pairs tend to depart from planarity in opposite directions, but these correlations are less striking than the positive ones. Taken together, the entries in Table 12 show clearly that Bchls 1, 2 and 3 correlate strongly with each other and therefore depart from planarity in the same way. We call this Class I. Similarly, Bchls 4, 5, 6 and 7 form a second distinct, highly correlated set (Class II). There is little if any (perhaps slightly negative) correlation between the 1, 2, 3 set and the 4, 5, 6, 7 set. As a further test we have included in Table 12 the correlations between Bchls 1-7 and the corresponding 22 atoms in the structure of ethylchlorophyllide (Chow et. al., 1975; Kratky & Dunitz, 1975). Clearly, the non-planarity of ethylchlorophyllide corresponds to the Bchl 4, 5, 6, 7 class and not to the Bchl 1, 2, 3 class. The high positive correlations between the non-planarity of ethylchlorophyllide and Bchls 4, 5 and 6, in particular, confirm that these deviations are physically meaningful. In particular, the deviations cannot be attributed to the restraints that were placed on the Bchl head-groups during refinement since no such restraints were applied to ethylchlorophyllide.

Figure 12 illustrates the out-of-plane distances averaged over the Class I and Class II Bchls. With the exception of the keto group, the

TABLE 12. Correlations Between Non-planarity of the Conjugated Bchl Macrocycles

		Class I			Class II				
	Bchl	1	2	3	4	5	6	7	E.C[(a)]
	1	1.0	0.56	0.32	-0.43	-0.37	-0.36	0.01	-0.35
Class I	2		1.0	0.84	-0.02	-0.12	-0.22	0.23	-0.19
	3			1.0	0.01	-0.15	-0.40	0.03	-0.36
	4				1.0	0.58	0.72	0.50	0.76
	5					1.0	0.80	0.46	0.80
Class II	6						1.0	0.62	0.92
	7							1.0	0.59
	E.C.								1.0

[(a)]E.C. is ethylchlorophillide; coordinates from Kratky & Dunitz (1975).

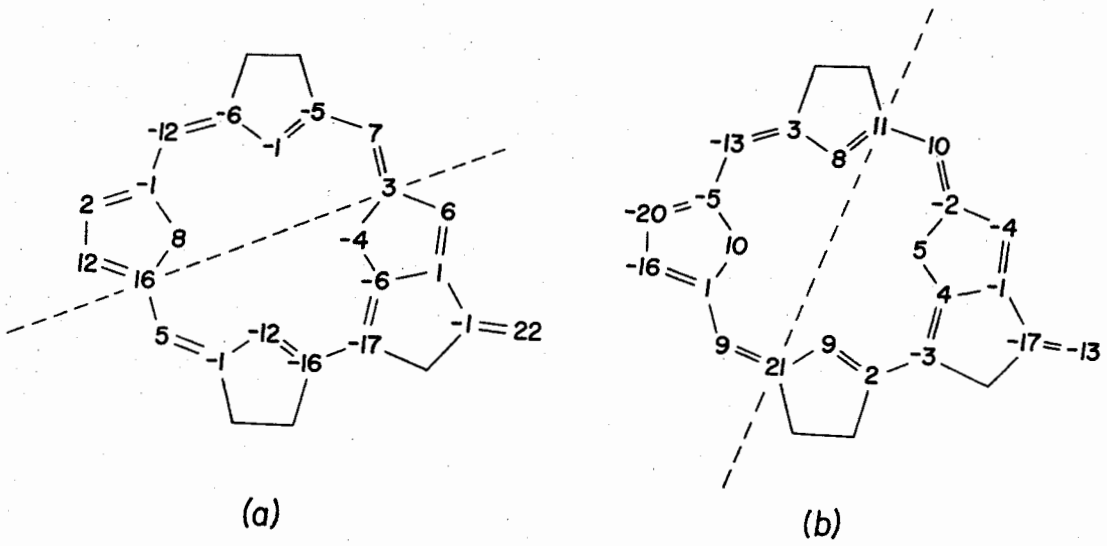


FIGURE 12. Schematic Drawing Illustrating the Out-of-Plane Distances of the Atoms in the Conjugated Chlorin System

The Bchls are in the same alignment as in Figure 9 and the numbers give the out-of-plane distances in 1/100ths of an angstrom unit. (a) Average distances for conjugated atoms in Bchls 1, 2 and 3. (b) Average distances for conjugated atoms in Bchls 4, 5, 6 and 7.

Class I deformation appears to be, in large part, a bending or curvature of the conjugated atoms about a line through atoms C1B and C1D. For the Class II Bchls there also seems to be overall bending or curvature of the conjugated macrocycle, in this case through a line connecting atoms C4A and C4C. Figure 13 shows the Class I and Class II Bchls superimposed in stereo.

Phytol Conformations

The phytyl chains of the seven Bchls and their interactions with the Bchl protein provide a model for lipid-protein interactions. Inspection of the phytols and their environments suggests the following general features (see Matthews, 1982 for a more detailed discussion).

Lipid Conformations Are Well-Defined But Irregular

In contrast to the fluid-like flexibility envisaged for lipids in a bilayer, the phytyl chains in the Bchl protein occupy well-defined positions in space with thermal factors comparable to the surrounding protein. In some cases the hydrocarbon chains are quite extended but they are not in idealized all-trans conformations. In this respect the Bchl phytols also differ markedly to crystal structures of isolated lipids in which very regular extended conformations are the rule (Pascher, 1976; Pascher & Sondell, 1977; Pearson & Pascher, 1979).

Lipid Tails Prefer Parallel Interactions

In a membrane bilayer, the amphipathic nature of membrane lipids results in the well-known bilayer structure with polar head-groups at

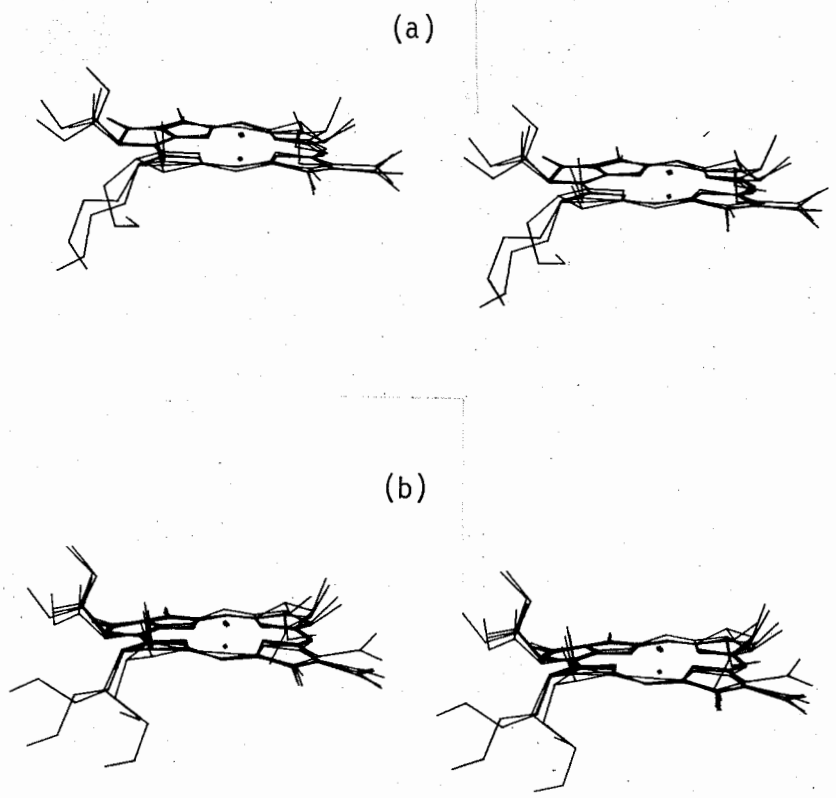


FIGURE 13. Superimposed Stereo Drawings of the "Class I" and "Class II" Bchl Head Groups

(a) Bchls 1, 2 and 3 (Class I). (b) Bchls 4, 5, 6 and 7 (Class II).

the surface and non-polar tails below the surface, interacting with each other. Because of these constraints the lipid tails are forced to lie more or less parallel with each other. In the Bchl protein, the constraints on the phytol packing is very different to that in a membrane bilayer, but where the hydrocarbon tails do interact they clearly prefer to do so in a parallel manner. There are seven cases where two phytyl chains from different Bchls remain in contact for five or more contiguous carbon atoms and in six of the seven cases the phytyl chains run in the same sense. The ester linkage to the chlorin head-group gives each phytol a distinct polarity, and, in addition, the asymmetric centers at PC8 and PC13 also imply polarity within the phytyl chain itself. Perhaps this inherent asymmetry favors parallel rather than antiparallel packing of the hydrocarbon chains.

Bends Occur At Substituted Positions

The individual phytyl chains have unique conformations and often make sharp bends. When such bends occur they are invariably at, or adjacent to, substituted positions (including CAA and CBA, between the chlorin ring and the phytol ester linkage). This is not to suggest that the hydrocarbon chains are fully extended except at or near substituted positions. Rather, there are local departures from the all-trans conformation but gross changes in direction occur at or near the sites of substitution.

Lipid-Lipid Interactions Do Not Dominate Over Lipid-Protein Interactions

Although there are extensive interactions between the phytyl chains of neighboring Bchls, these interactions do not dominate over interactions between the phytols and the surrounding protein. Inspection of the structure suggests that there is roughly the same contact area occupied in phytol-phytol, phytol-chlorin and phytol-protein interactions.

CHAPTER V

CRYSTALLOGRAPHIC STRUCTURAL ANALYSIS OF PHOSPHORAMIDATES AS INHIBITORS AND TRANSITION STATE ANALOGS OF THERMOLYSIN

Abstract

The mode of binding to thermolysin of the unsubstituted phosphoramidate inhibitor N-phosphoryl-L-leucinamide (P-Leu-NH_2) has been determined crystallographically and refined at high resolution (R = 17.9% to 1.6Å resolution). The mode of binding of the naturally occurring thermolysin inhibitor phosphoramidon reported previously (Weaver et al., 1977) has also been confirmed by crystallographic refinement (R = 17.4% to 2.3Å resolution). Phosphoramidon binds to the enzyme with a single oxygen of the phosphoramidate moiety as a zinc ligand. Together with three ligands to the metal from the protein the resultant complex has approximately tetrahedral geometry. However, in the case of P-Leu-NH_2, two of the phosphoramidate oxygens interact with the zinc to form a complex that tends towards pentacoordinate. In this respect, P-Leu-NH_2 appears to be a better transition state analog than is phosphoramidon. In addition, the phosphorous-nitrogen bond length in P-Leu-NH_2 is 1.8Å, suggesting that the nitrogen is protonated whereas the same bond in phosphoramidon is much shorter (1.5Å) suggesting that the nitrogen does not carry a charge. In phosphoramidon the distance from the phosphoramide nitrogen to Glu 143 is 3.9Å whereas in P-Leu-NH_2

this distance decreases to 3.4Å. Taken together, these observations provide additional evidence in support of the participation of pentacoordinate intermediates in the mechanism of action of thermolysin (Holmes & Matthews, 1981) and the role of Glu 143 in first promoting the attack of a water molecule on the carbonyl carbon of the scissile bond and subsequently acting as a "proton shuttle" to transfer the proton to the leaving nitrogen (Monzingo & Matthews, 1984; Hangauer et al., 1984).

Introduction

Thermolysin [EC 3.4.24.4] is a thermostable zinc endopeptidase from the thermophilic bacterium Bacillus thermoproteolyticus. Although the amino acid sequence of thermolysin is not related to the sequences of other zinc-requiring peptidases such as carboxypeptidase A and the angiotensin converting enzyme, there is increasing evidence that the active sites of these zinc enzymes have features in common (Kester & Matthews, 1977b; Cushman et al., 1977; Ondetti et al., 1977, 1979; Kam et al., 1979; Holmquist & Valee, 1979; Maycock et al., 1981).

The structure of thermolysin is known to high resolution (Holmes & Matthews, 1982) and the modes of binding of a variety of inhibitors including dipeptides, mercaptans, hydroxamates and N-carboxymethyl peptides have been determined (e.g. see Holmes & Matthews, 1981; Kester & Matthews, 1977a; Monzingo & Matthews, 1983). These studies, together with a computer graphics analysis of the thermolysin active site have led to a detailed stereochemical proposal for the mechanism of action of thermolysin (Hangauer et al., 1984).

In a previous paper we determined the mode of binding to thermolysin of the naturally occurring inhibitor phosphoramidon (Weaver et al., 1977). In the present report we present the refined coordinates for phosphoramidon and, in addition, describe the mode of binding to thermolysin of N-phosphoryl-L-leucinamide. The previous analysis of phosphoramidon, a substituted phosphoramidate, indicated that this naturally occurring inhibitor binds to thermolysin with monodentate zinc ligation (Weaver et al., 1977). This result is confirmed by crystallographic refinement. However, the unsubstituted phosphoramidate is here shown to bind to the zinc with coordination tending toward bidentate, i.e. toward the coordination of the transition state suggested by the computer graphics analysis.

Phosphoramidon (N-(α-L-rhamnopyranosyl-oxyhydroxyphosphinyl)-L-leucyl-L-tryptophan) is a naturally occurring potent inhibitor of thermolysin (K_I = 28 nM) (Suda et al., 1973; Komiyama, 1975). Based on the prior crystallographic analysis it was proposed as an analog for the presumed tetrahedral intermediate formed during the hydrolysis of peptides (Weaver et al., 1977). Also based on the effectiveness of phosphoramidon as an inhibitor of thermolysin, a series of related phosphoramidates and phosphonamidates have been synthesized and have proved to be potent inhibitors not only of thermolysin, but also of other zinc peptidases including carboxypeptidase A and the angiotensin converting enzyme (Bartlett & Marlowe, 1983; Galardy, 1982; Galardy et al., 1983; Holmquist, 1977; Holmquist & Vallee, 1979; Jacobsen & Bartlett, 1981; Kam et al., 1979; Komiyama et al., 1975; Nishino & Powers, 1979;

Thorsett et al., 1982). These studies clearly show the importance of the tetrahedral N-phosphoryl moiety in the binding of these inhibitors. For example, the simple phosphoramidate N-phosphoryl-L-leucinamide (P-Leu-NH_2) is an excellent inhibitor of thermolysin (K_I = 1.3 μM) (Kam et al., 1979). In general, the addition of substituents to the phosphoramidate group tends to decrease the effectiveness of the inhibitor. Methylation, for example, as in O-methyl-phosphoryl-Leu-NH_2 weakens the binding by about 100-fold relative to P-Leu-NH_2, and O-methyl-O-phenylphosphoryl-Leu-NH_2 does not inhibit thermolysin (Kam et al., 1979). Likewise, removal of the rhamnose from phosphoramidon itself results in a slightly tighter-binding inhibitor (Kam et al., 1979; Komiyama et al., 1975). Recently, Bartlett and Marlowe (1983) have synthesized a series of six phosphorous-containing peptide analogs and provided evidence that these phosphonamidate inhibitors of thermolysin are transition state analogs and not just multi-substrate ground state analogs.

If, indeed, phosphoramidates and phosphonamidates can be considered as transition state analogs, then it is of interest to determine their geometry of binding as precisely as possible. In this paper we present the high resolution crystallographic analysis of the binding to thermolysin of the unsubstituted phosphoramidate N-phosphoryl-L-leucinamide (P-Leu-NH_2) (a kind gift of Dr. J.C. Powers). Subsequent reports will describe the modes of binding of additional phosphorous-containing peptide and ester analogues developed by P.A. Bartlett and coworkers (cf. Bartlett & Marlowe, 1983).

Materials and Methods

Thermolysin, obtained from Calbiochem, was crystallized as described by Holmes and Matthews (1982). Native crystals were equilibriated with a mother liquor containing 0.01 M calcium acetate, 0.01 M Tris-acetate, 7% dimethylsulfoxide (V/V), pH 7.2. P-Leu-NH_2 was dissolved in the mother liquor and then added to a vial containing native crystals. Binding of the inhibitor to crystalline thermolysin was monitored by calculating (hOℓ) difference Fourier projections from precession photographs (cf. Weaver et al., 1977).

One crystal soaked for 18 days with 4×10^{-4} M P-Leu-NH_2 was used for three-dimensional data collection. Data were measured to 1.6Å resolution using oscillation photography (Rossmann, 1979; Schmid et al., 1981). An oscillation angle of 1.2° per film pack was used and the crystal, space group $P6_122$, was rotated about the c axis through a net rotation of 30°. Approximately half the data set were collected by exposing one end of the crystal (a long hexagonal rod). The crystal was then translated along the spindle axis to the other end, and the remainder of the data were collected. The X-ray source was a graphite-monochromatized Elliot GX-21 rotating anode generator operated at 39 kV, 130 mA, and the exposure time was 6 hours per film pack. Three films were used in each pack. Intensity statistics are summarized in Table 13.

Model-Building of P-Leu-NH_2 Bound to Thermolysin

A difference electron density map was calculated with amplitudes ($F_{complex} - F_{Nat}$) and with phases calculated from the refined native

TABLE 13. Intensity Statistics for P-Leu-NH_2

Films	28
Average $R_{sym}^{(a)}$	0.032
Average $R_{sca}^{(a)}$	0.028
$R_{merge}^{(a)}$	0.050
Average isomorphous difference (%)	13.2
Independent reflections	31,621

(a) $R = \Sigma \left| I-\bar{I} \right| \Sigma I$. R_{sym} measures the agreement between symmetry-related reflections on the same film, R_{sca} measures the agreement between reflections recorded on successive films in a given film pack and R_{merge} gives the overall agreement between intensities measured on different films.

structure (Holmes & Matthews, 1982). The largest peak of the map was close to the zinc ion and was presumed to represent the density of the phosphorous atom of the inhibitor (Fig. 14). The height of this peak is 27σ where σ is the root-mean-square density of the difference map through the unit cell. For detailed model building, an electron density map (Fig. 15) was calculated with amplitudes of the form $(2F_{Complex} - F_{Nat.,calc})$, the native structure factors and phases being calculated from the refined native coordinates with active site water molecules removed (c.f. Holmes & Matthews, 1981).

Refinement of the Thermolysin:P-Leu-NH_2 Complex

The atomic coordinates of the thermolysin:P-Leu-NH_2 complex, including bound water molecules, were partially refined using the energy minimization and crystallographic refinement program of Jack and Levitt (1971), modified by J. Deisenhofer and by D.H. Ohlendorf. The refinement was completed using a restrained least squares method developed in this laboratory by D.E.T. and L.F. Ten Eyck (see Chapter III). The refinement package ("TNT") is both general and flexible and can readily accommodate the "non-standard" geometry of this inhibitor. The final crystallographic R-value was 17.9% to 1.6Å resolution. Refined coordinates of the inhibitor, listed in Table 14, have an estimated uncertainty of not more than 0.15Å (cf. Holmes & Matthews, 1982). Refinement statistics are summarized in Table 15. The deviations from "ideal" geometry for the refined protein-inhibitor complex and the inhibitor alone are shown in Table 16.

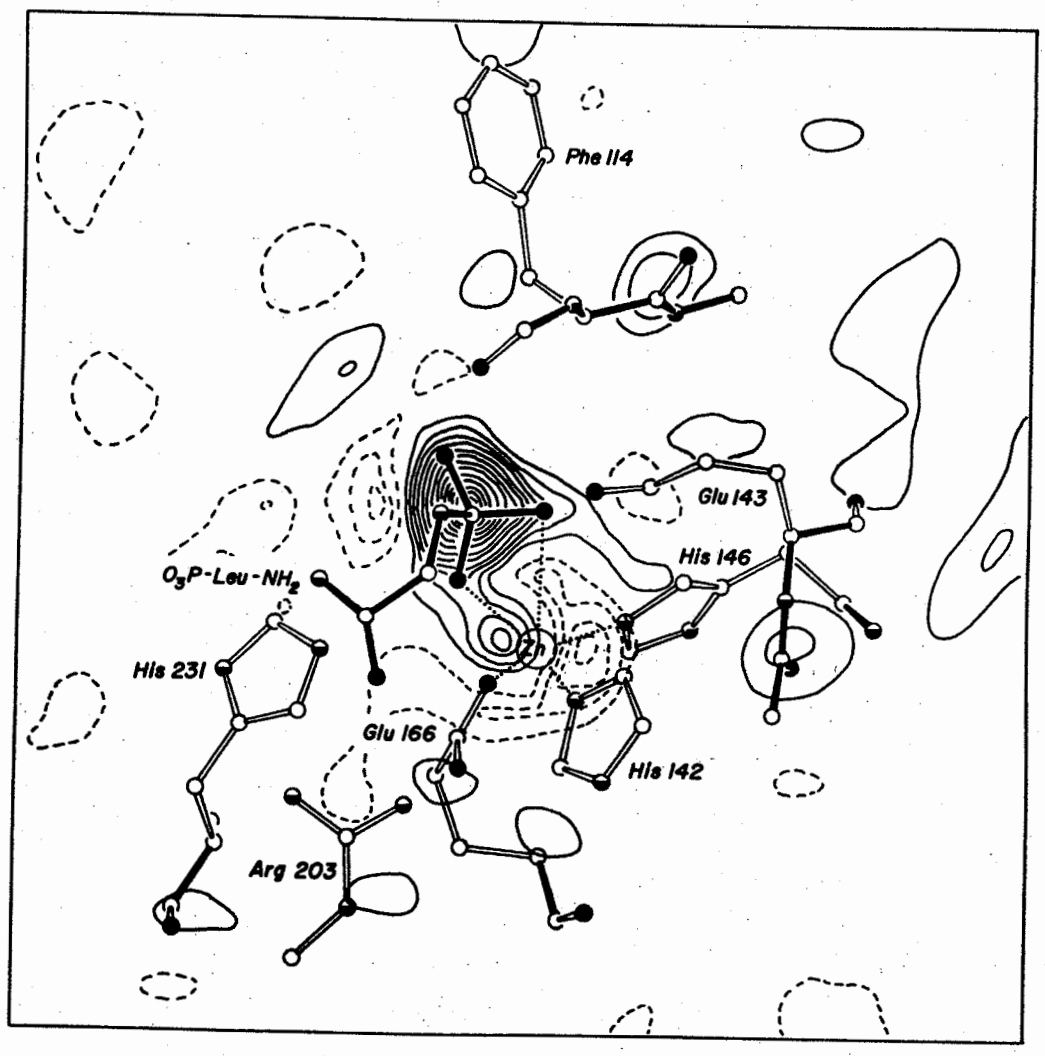


FIGURE 14. Section of Difference Map Cutting Through the Bound Inhibitor of P-Leu-NH_2

Section $\underline{z}$ = -0.05 of the 1.9Å resolution electron density difference map with amplitudes ($F_{Complex} - F_{Nat.obs}$) and phases calculated from the refined native structure. Contours are drawn at intervals of 2.0σ, with positive contours drawn solid and negative contours broken. Protein and inhibitor atoms lying close to this section are included; oxygen atoms are drawn solid, nitrogen half solid and carbon as open circles. Inhibitor bonds and protein backbone bonds are drawn solid.

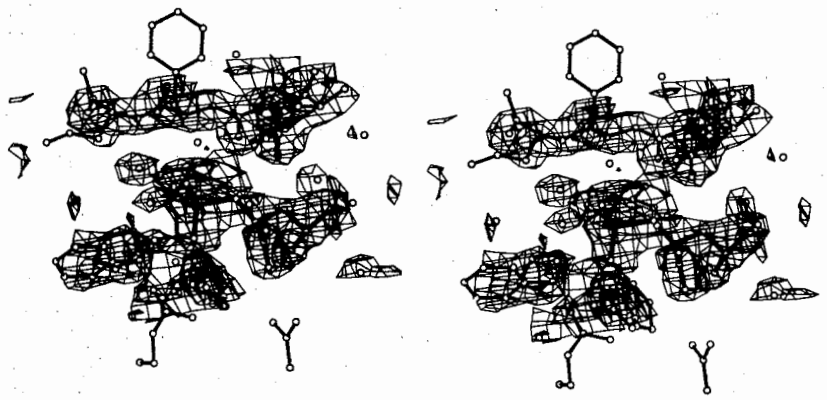


FIGURE 15. Stereo Drawing of Difference Map for Inhibitor P-Leu-NH_2

Difference electron density map, resolution 1.9Å, with coefficients ($F_{complex}$ - $F_{Nat.,calc}$) superimposed on the model for the bound inhibitor (shown with thick bonds). The direction of view is approximately the same as in Figure 16. The native amplitudes and phases were calculated from the refined native structure with active-site solvent molecules removed. In several cases the protein atoms extend beyond the volume of contoured electron density.

TABLE 14. P-Leu-NH_2 Coordinates

Atom	X	Y	Z	B
PO3	50.6	18.3	-7.0	14
PO2	53.2	18.5	-6.8	12
PO1	51.7	20.5	-6.2	17
P	51.7	19.0	-6.3	15
N	51.6	18.4	-4.5	13
CB	52.5	18.3	-2.2	11
CG	53.7	18.2	-1.3	17
CD1	54.6	19.5	-1.5	15
CD2	52.9	18.2	0.0	14
CA	52.8	18.3	-3.7	17
C	53.7	17.0	-4.0	17
O	54.9	17.1	-3.9	14
NH2	53.0	15.9	-4.3	14
Sol1:0362	49.9	21.2	-8.0	18

Coordinates are in Angstroms in the standard orthogonal thermolysin coordinate system (Weaver et al., 1974). B is the crystallographic thermal parameter in $Å^2$.

TABLE 15. Refinement Statistics for Thermolysin:Inhibitor Complexes

	P-Leu-NH_2	Phosphoramidon
Resolution	1.6Å	2.3Å
Initial R-factor	23.3%	22.1%
Final R-factor	17.9%	17.4%
Number of cycles	21	16
Number of reflections used	31,416	13,523
Number of atoms	2,618	2,637

TABLE 16. Weighted Deviations From "Ideal" Geometry for Refined Thermolysin:Inhibitor Complexes

	P-Leu-NH_2	Phosphoramidon
	Rms deviation	Rms deviation
Bond length	0.023Å	0.022Å
Bond angle	3.6°	2.9°
Planarity (trigonal)	0.018Å	0.026Å
Planarity (other planes)	0.018Å	0.016Å

Refinement of the Thermolysin:Phosphoramidon Complex

The previously reported structure of the complex of phosphoramidon with thermolysin was based on stereochemically idealized, but otherwise unrefined coordinates (Weaver et al., 1977). With these coordinates as a starting point, the complex was refined with the "TNT" package to a crystallographic R-value of 17.4% at a nominal resolution of 2.3Å (Tables 15, 16). The refined coordinates for phosphoramidon are given in Table 17; their estimated accuracy is about 0.25Å.

Results

Binding of Phosphoramidon

During refinement, the phosphoramidon coordinates changed only slightly (root-mean-square shift of 0.25Å). The previous description of the binding of this inhibitor to crystalline thermolysin (Weaver et al., 1977) is therefore confirmed by the refinement. Details of the interactions between phosphoramidon and thermolysin, based on the refined coordinates, are included in Tables 18 and 19. Refined coordinates for both the thermolysin:phosphoramidon complex and the P-Leu-NH_2 complex have been deposited in the Brookhaven Data Bank (Bernstein et al., 1977).

Binding of P-Leu-NH_2

The observed binding of P-Leu-NH_2 to thermolysin is shown in Figure 16. The zinc is coordinated by one phosphoryl oxygen (PO1) of the inhibitor at a distance of 2.1Å. A second phosphoryl oxygen (PO2) is

TABLE 17. Phosphoramidon Coordinates

Atom	X	Y	Z	B
Rhamnose				
C1	51.2	16.9	-8.3	21
O1	50.9	18.1	-7.6	22
C2	50.2	16.1	-7.5	23
O2	50.0	14.7	-8.0	28
C3	48.9	16.8	-7.8	22
O3	48.0	16.0	-6.9	36
C4	48.6	16.5	-9.3	24
O4	47.3	16.7	-9.6	50
C5	49.6	17.4	-10.1	18
O5	50.9	17.0	-9.7	30
C6	49.4	17.4	-11.6	32
Phosphoramide				
P	51.7	18.9	-6.6	17
PO1	53.2	19.1	-6.9	5
PO2	51.2	20.2	-6.3	15
N	51.7	18.1	-5.4	24
Leucine				
CA	52.7	18.0	-4.4	14
C	53.5	16.7	-4.4	11
O	54.6	16.9	-3.9	6

TABLE 17. (continued)

CB	52.1	18.2	-3.0	18
CG	53.2	18.1	-1.9	24
CD1	54.4	19.1	-1.8	21
CD2	52.6	18.1	-0.4	13
Tryptophan				
N	52.9	15.6	-4.8	18
CA	53.5	14.2	-5.0	22
C	52.7	13.2	-5.8	21
O	51.5	13.6	-5.9	18
CB	54.0	13.8	-3.6	16
CG	52.9	13.4	-2.7	20
CD1	52.2	14.2	-1.9	21
NE1	51.2	13.5	-1.3	18
CE2	51.4	12.2	-1.5	18
CZ2	50.7	11.1	-1.1	33
CH2	51.1	9.8	-1.5	31
CZ3	52.1	9.7	-2.4	34
CE3	52.9	10.8	-2.9	31
CD2	52.5	12.0	-2.4	20
OH	53.2	12.1	-6.2	40
Solvent molecule				
O1	49.8	21.4	-8.0	10

TABLE 18. Zinc-Ligand Geometry

Ligand	Distance in P-Leu-NH_2 Complex (Å)	Distance in Phorphoramidon Complex (Å)
His 142 NE2	2.1	2.2
His 146 NE2	2.1	2.2
Glu 166 OE1	2.2	2.0
Inhibitor PO1	2.1	1.8
Inhibitor PO2	2.8	3.4
(Tyr 157)	3.7	3.7

Ligands	Angle in P-Leu-NH_2 Complex	Angle in Phosphoramidon Complex
PO1-Zn-NE2 (His 142)	119.0°	116°
PO1-Zn-NE2 (His 146)	128.0°	114°
PO1-Zn-OE1 (Glu 166)	92.4°	99°

Table 19. Selected Thermolysin-Inhibitor Distances[a]

Protein	Inhibitor	Distance in P-Leu-NH_2 Complex	Distance in Phosphoramidon Complex
		(Å)	(Å)
Tyr 157 OH	O5	-	2.8
Tyr 157 OH	PO1	3.2(H)	3.2
Glu 166 OE1(Zn)	PO1	3.0	2.8
His 231 NE2	PO1	2.8(H)	3.2(H)
His 142 NE2(Zn)	PO2	3.7	4.5
His 146 NE2(Zn)	PO2	3.2	3.6
Glu 143 OE1	PO2	2.4(H)	2.5(H)
Glu 143 OE2	PO2	3.0	3.2
Asn 112 OD1	Leu N	3.1	3.4
Ala 113 O	Leu N	3.0(H)	3.4
Glu 143 OE2	Leu N	3.4	3.9
His 231 NE2	Leu N	4.6	4.0
Arg 203 NH1	Leu O	2.9(H)	3.0(H)
Arg 203 NH2	Leu O	2.8(H)	3.1(H)
Asn 112 OD1	Amide or Trp N	3.5	3.7
His 231 NE2	Amide or Trp N	3.4	3.2
Asn 112 ND2	Trp O	-	3.2(H)
Ala 111 O	Trp NE1	-	3.0(H)

TABLE 19. (continued)

Distance to solvent molecule			
Trp 115 N		2.9(H)	2.9(H)
Glu 143 OE1		3.3	3.2(H)
	PO2	2.7(H)	2.5(H)
	PO3	3.1(H)	3.5

(a) (H) indicates a presumed hydrogen bond; (Zn) indicates a zinc ligand.

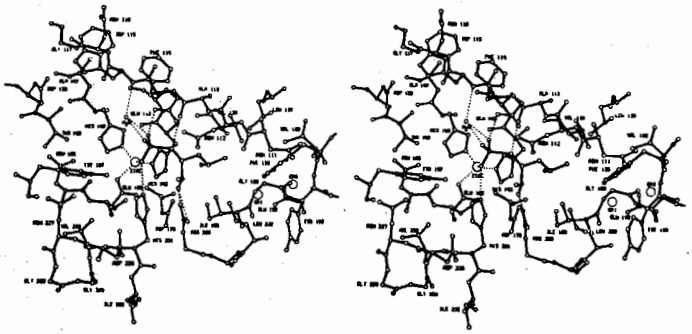


FIGURE 16. Model for P-Leu-NH_2 Shown in the Extended Thermolysin Active Site

The inhibitor and the main-chain of the protein are drawn with solid bonds; protein side chains are drawn with open bonds. Apparent hydrogen bonds between the inhibitor and the protein are shown as dotted lines.

2.8Å from the zinc, and can be considered as a weak ligand. The interatomic distances between zinc and all of its ligands are listed in Table 18. The leucyl side chain of the inhibitor is nestled in the hydrophobic specificity pocket as has been observed with all good thermolysin inhibitors (e.g. see Holmes & Matthews, 1981; Monzingo & Matthews, 1983; Weaver et al., 1977).

All close interatomic contacts between the protein and the inhibitor are included in Table 19. In addition to its coordination to the zinc, PO1 is 2.8Å from the NE2 atom of His 231 and PO2 is within hydrogen bonding distance of both the carboxylate oxygens of Glu 143. The potential significance of these and other interactions in the vicinity of the zinc are discussed in more detail in the next section. As seen with other thermolysin inhibitors, the carbonyl oxygen of the R_1' residue is within hydrogen-bonding distance of both the NH1 atom (2.9Å) and the NH2 atom (2.8Å) of Arg 203.

Discussion

The crystallographic refinement of the thermolysin:phosphoramidon complex confirms the description of this structure given previously (Weaver et al., 1977). In general terms, the structure of the complex of thermolysin with P-Leu-NH_2 resembles that with phosphoramidon, although there are significant differences. For comparison, the structures of P-Leu-NH_2 and phosphoramidon are shown superimposed in the thermolysin active site (Fig. 17). The respective locations of the leucyl side chains in the S_1' specificity pocket are similar as are the

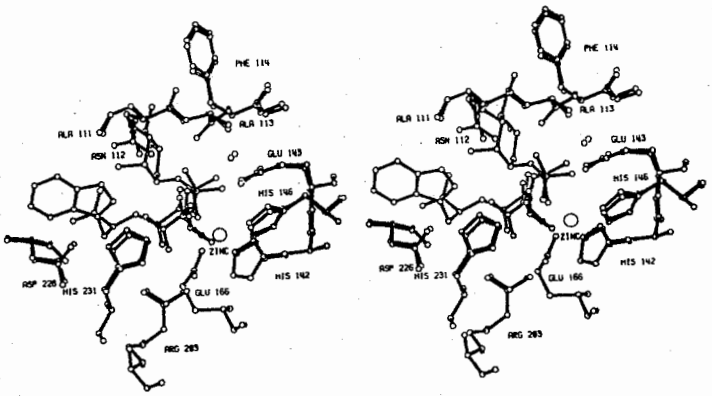


FIGURE 17. Comparison of Binding of P-Leu-NH_2 and Phosphoramidon

Superposition of P-Leu-NH_2 (solid bonds) on phosphoramidon (open bonds) bound within the thermolysin active site.

the hydrogen bonds between the R_1' carbonyl oxygen and Arg 203, and the R_1' amino nitrogen and the carbonyl oxygen of Ala 113. These interactions are all characteristic of a variety of inhibitors of thermolysin (e.g. see Holmes & Matthews, 1981; Kester & Matthews, 1977a). However, the two inhibitors differ significantly in the placement of their phosphoramidate groups. In P-Leu-NH_2, the phosphoramidate oxygens are rotated by about 25° relative to phosphoramidon (Fig. 17). The nitrogen of the phosphoramidate also differs by about 0.9Å in the two inhibitors.

We previously suggested that the tetrahedral phosphoramidate group is an analog for the tetrahedral intermediate formed by the presumptive attack of a water molecule or hydroxide ion on the carbonyl carbon of the scissile bond (Weaver et al., 1977). It was subsequently observed that hydroxamic acid inhibitors bound to thermolysin with bidentate zinc ligation, suggesting that intermediates with overall five-fold zinc coordination might occur during catalysis (Holmes & Matthews, 1981. Independently, Kunugi et al. (1982) argued for five-fold zinc coordination based on the pH and temperature dependence of thermolysin catalysis. There is also spectroscopic evidence for pentacoordinate intermediates in the case of carboxypeptidase A (Kuo & Makinen, 1982).

In a recent analysis, the mode of binding to thermolysin of an N-carboxymethyl dipeptide inhibitor was determined crystallographically (Monzingo & Matthews, 1983). This inhibitor was shown to bind with the carboxyl of the N-carboxymethyl group forming a bidentate complex with the zinc (to give overall 5-fold coordination) and the nitrogen of the N-carboxymethyl group close to Glu 143. Based on the above findings a

computer graphics analysis was used to develop a detailed stereochemical proposal for the mechanism of action of thermolysin (Hangauer et al., 1984). The essential features of the mechanism are as follows: Native thermolysin has a single water molecule liganded to the zinc. Together with three ligands from the protein, the zinc coordination is essentially tetrahedral. As a peptide substrate binds, the carbonyl carbon forms a fifth zinc ligand and displaces the bound water molecule towards Glu 143. The water molecule, activated by the combined influence of Glu 143 and the metal, attacks the carbonyl carbon of the substrate forming a tetrahedral transition state intermediate in which two of the oxygens are liganded to the zinc. The proton, accepted by Glu 143 from the attacking water molecule, is shuttled to the leaving nitrogen, leading to cleavage of the scissile bond and displacement of products (Hangauer et al., 1983; Monzingo & Matthews, 1983). In such a mechanism Glu 143 performs the dual role of assisting the attack of the water molecule on the carbonyl carbon of the peptide bond and also acting as the proton donor to the nitrogen. By analogy, with thermolysin, a similar mechanism was proposed for carboxypeptidase A (Monzingo & Matthews, 1983) and has been strongly supported by the application of site-directed mutagenesis (Gardell et al., 1985).

For the crystallographic complex of thermolysin and phosphoramidon, the zinc ligation was seen to be approximately tetrahedral, with a single oxygen ligand donated by the inhibitor. This result is confirmed by the present refinement (Table 17). In the case of phosphoramidon, the alignment of the phosphoramidate group is constrained by the

presence of the rhamnose. However, for P-Leu-NH_2, there are no substituents (excepting protons) on PO1, PO2 or PO3, so that this group can presumably rotate freely to allow the three oxygens to maximize their interactions with the enzyme. Therefore the complex of P-Leu-NH_2 with thermolysin ought to illustrate the preferred configuration of the free phosphoramidate group, and the interactions that stabilize it. (Of course this need not be the same configuration as adopted by a (substituted) tetrahedral intermediate during catalysis.)

Indeed the phosphoramidate group of P-Leu-NH_2 does rotate relative to that in phosphoramidon (Fig. 17), and this motion, coupled with small changes in the protein, results in significant differences in several of the protein:inhibitor contacts (Fig. 18; Table 19). (In comparing the two sets of distances in Tables 18 and 19 it should be remembered that the P-Leu-NH_2 data are to higher resolution (1.6Å) than phosphoramidon (2.3Å) so that the quoted distances are more reliable for P-Leu-NH_2.) In P-Leu-NH_2, oxygen PO2 is 2.8Å from the zinc, 0.6Å closer than for phosphoramidon, and can be considered as a weak ligand. The zinc ligation is substantially distorted from tetrahedral (Table 18) and tends toward pentacoordinate, although not to the extent seen for hydroxamates (Holmes & Matthews, 1981) and for the N-carboxymethyl inhibitor described by Monzingo and Matthews (1983). For P-Leu-NH_2, oxygen PO2 remains within hydrogen bonding distance of Glu 143, as in phosphoramidon, but seems to be better placed to hydrogen bond to either, or both, of the carboxylate oxygens. Oxygen PO1 is the principal zinc ligand for both inhibitors, and, in the case of

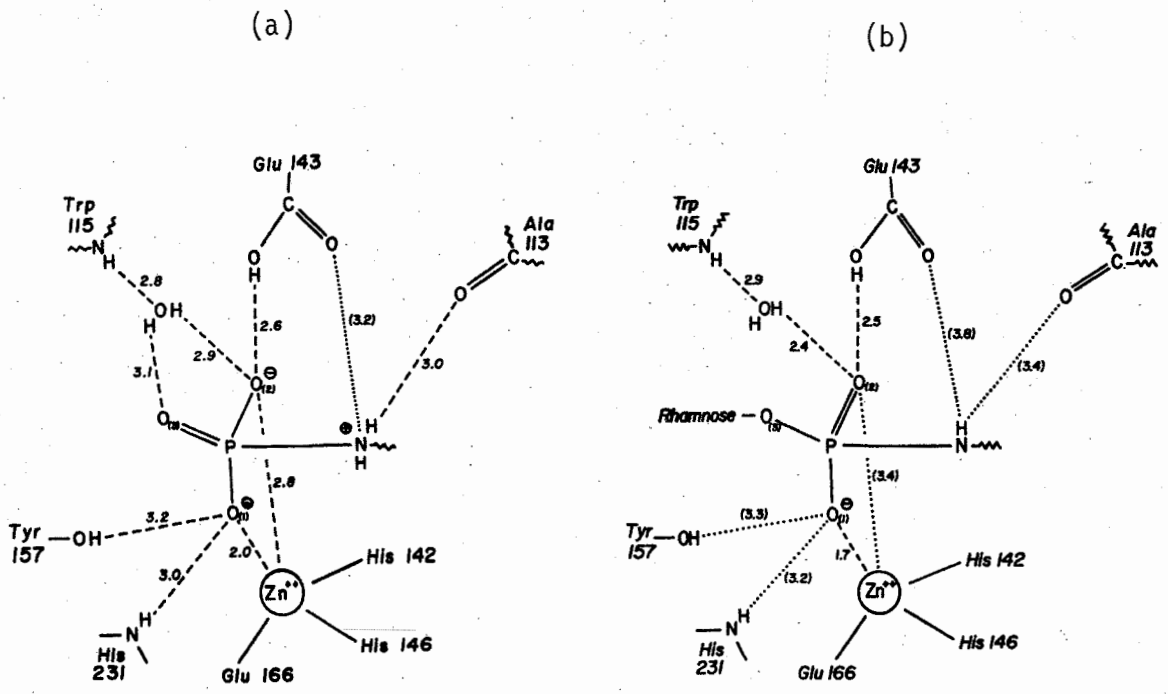


FIGURE 18. Simplified Sketch Showing the Apparent Interactions Between Two Phosphoramidate Inhibitors and Thermolysin

Presumed hydrogen bonds and interactions with the zinc are drawn as broken lines, other close approaches are drawn as dotted lines with the distances indicated in parentheses. (a) Interactions for P-Leu-NH_2. (b) Interactions for phosphoramidon.

P-Leu-NH_2 at least, seems to be hydrogen bonded by nitrogen NE2 of His 231. For phosphoramidon, the NE2-PO1 distance of 3.2Å is rather long for a hydrogen bond, but the shorter distance of 2.8Å observed for P-Leu-NH_2 tends to support such an interaction.

There is also a difference between the probable ionization state of phosphoramidon and P-Leu-NH_2 that needs to be considered (Kam et al., 1979). Phosphoramidon should be ionized at neutral pH with the negatively charged oxygen thus formed expected to coordinate to the electropositive zinc. However, at neutral pH, phosphoramidates are expected to exist as zwitterionic monoanions, with the nitrogen protonated and the PO_3 group carrying two negative charges (Jacobsen & Bartlett, 1981; Kam et al., 1979). The crystallographic refinement provides independent evidence for this difference in protonation. In the refined thermolysin:phosphoramidon complex, the length of the phosphorous-nitrogen bond is 1.5Å whereas in P-Leu-NH_2 this bond length increases to 1.8Å. The difference in length is consistent with a change from an essentially double bond in phosphoramidon to a single bond in P-Leu-NH_2. This, in turn, is indicative of a change from an unprotonated (trigonal) nitrogen in phosphoramidon to a protonated (tetrahedral) nitrogen in P-Leu-NH_2. (In the respective crystallographic refinements the P-N bond length was weakly restrained to an "expected" value of 1.78Å. σ for the P-N bond was made three times as large as for all other bonds. This corresponds to a relative weight of 1/9 in the least squares refinement.) Although the estimated errors in the coordinates of the phosphorous and nitrogen atoms are about 0.15-0.2Å, these errors are

correlated, and it is not easy to estimate the likely error in the P-N bond. For this reason we carried out several controls to test whether the apparent difference in the P-N bond length in phosphoramidon and P-Leu-NH_2 is real. For example, we lengthened the P-N bond length in phosphoramidon to 1.78Å and re-refined the structure. The bond length still reverts back to a value close to 1.5Å.

Figure 18 shows possible resonance structures for phosphoramidon and P-Leu-NH_2 complexed to thermolysin. Although the pH of the thermolysin crystals is 7.2 we have nevertheless drawn Glu 143 in Figure 18 assuming it to be protonated. This is done in part because the short distance between Glu 143 and oxygen PO1 (Figs. 18(a), 18(b)) indicates that this must be a hydrogen bond. In addition, Palmer et al. (1982) have shown that Glu 270 of carboxy-peptidase A is protonated when the negatively charged inhibitor benzylsuccinic acid is bound and we believe that this carboxypeptidase:inhibitor complex has close parallels with thermolysin (Bolognesi & Matthews, 1979).

Kam et al. (1979) showed that unsubstituted phosphoramidates bind to thermolysin more tightly than their monoalkyl esters and proposed that the enhanced stabilization in the former case might be due to additional interactions with Tyr 157 and His 231. Indeed, Tyr 157 and His 231 do appear to stabilize the binding of P-Leu-NH_2 (Fig. 18(a)) although the refined coordinates for phosphoramidon suggest that very similar interactions can occur for at least this example of a substituted phosphoramidate. Comparison of Figures 18(a) and 18(b) suggests three additional reasons why substitution of the

phosphoramidate might weaken binding. First, the unsubstituted phosporamidate can have a more favorable ionic interaction with the zinc, with two oxygens acting as zinc ligands. Second, the protonated nitrogen of the unsubstituted phosphoramidate appears to form a stronger hydrogen bond (3.0Å) to the carbonyl oxygen of Ala 113 than is the case with phosphoramidon (3.4Å). This change in geometry is presumably associated with the change of the nitrogen from non-protonated to protonated. It is noteworthy that the protonated form, as in the presumed transition state, is the one preferred. The third possible reason why an unsubstituted phosphoramidate might be bound more tightly than a substituted one is that the solvent structure might be more favorable. When both phosphoramidon and P-Leu-NH_2 bind to thermolysin, there is a water molecule that bridges between the inhibitor and the enzyme (Figs. 18(a), 18(b)). In the case of phosphoramidon this water molecule makes a single hydrogen bond to the inhibitor, namely to PO2 (Fig. 18(b)). However, in the case of P-Leu-NH_2, the water molecule makes hydrogen bonds to two oxygens of the inhibitor (Fig. 18(a)), in what is presumably an energetically favorable situation.

It might be asked why the phosphoramidate moiety of P-Leu-NH_2 does not form a fully bidentate complex with the two oxygens equidistant from the zinc. The congestion around the zinc, compounded by the size of the phosphoramide group, may prevent a completely symmetric mode of binding. In any event, the interactions of PO1 and PO2 with the protein and bound solvent are different (Fig. 18(a)) and when these are taken into account it would be surprising if the optimal configuration of the phosphoramide

group was one in which PO1 and PO2 were exactly equidistant from the metal.

One of the largest changes, and perhaps the most suggestive, in comparing P-Leu-NH_2 with phosphoramidon, involves the phosphoramidate nitrogen ("Leu N" in Table 19). In the P-Leu-NH_2 complex (Fig. 18(a)) this nitrogen is 0.4Å closer to Glu 143 (from 3.8Å to 3.4Å), and, at the same time, 0.5Å further away from His 231 (from 4.0 to 4.5Å), relative to phosphoramidon (Fig. 18(b)). As noted by Monzingo and Matthews (1983) the N-carboxymethyl dipeptide inhibitor described there also has its α-amino group, which is possibly protonated, close (3.0Å) to Glu 143, suggesting that Glu 143 should be considered as a potential proton donor. The relationship between the geometry of binding of P-Leu-NH_2 and the N-carboxymethyl dipeptide inhibitor is illustrated in Fig. 19. A detailed model-building analysis (Hangauer et al., 1984) is consistent with a mechanism of action in which Glu 143 accepts a proton from the attacking water molecule and subsequently shuttles this proton to the leaving nitrogen. The simultaneous close approach of Glu 143 to both oxygen PO2 and the amide nitrogen of P-Leu-NH_2 provides additional support for this mechanism. Also, the overall geometry of binding of the phosphoramide group is as good or better an approximation to the presumed transition state than other thermolysin inhibitors studied previously (Fig. 20). The model-building analysis (Hangauer et al., 1984) suggests that, in order to retain acceptable stereochemistry and to maximize favorable protein:substrate interaction, the tetrahedral group in the presumptive transition state is rotated even further than that

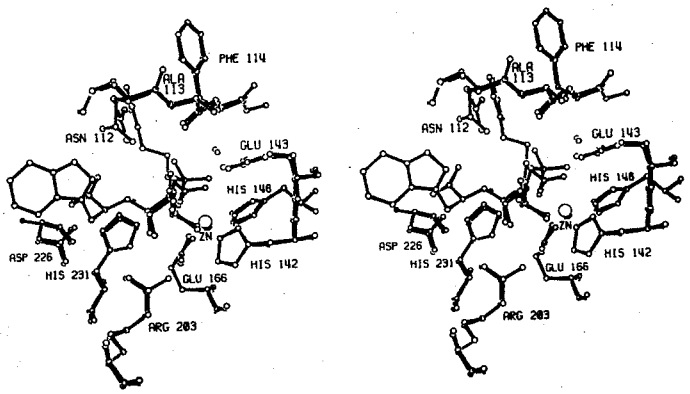


FIGURE 19. Comparison of Binding of P-Leu-NH_2 and CLT

Superposition of P-Leu-NH_2 (solid bonds) and N-(1-carboxy-3-phenylpropyl)-L-leucyl-L-tryptophan (CLT) bonds as these two inhibitors bind to thermolysin.

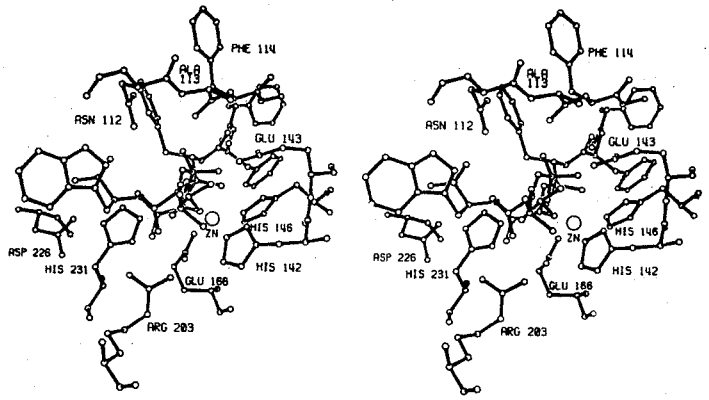


FIGURE 20. Comparison of Binding of P-Leu-NH_2 and a Model for the Transition State

Superposition of P-Leu-NH_2 (solid bonds) bound in the thermolysin active site on the model for the presumed transition state (open bonds) as proposed by Hangauer et al. (1984).

seen in the comparison of phosphoramidon with P-Leu-NH_2 (Figs. 17, 20). It is assumed that in the transition state, the oxygen corresponding to PO2 moves closer to the zinc to form a full bidentate complex.

While the model building study does not rule out the possibility that a proton is donated either by His 231 or a water molecule, this seems less likely than considered previously (Kester & Matthews, 1977a). In the complex of thermolysin with the N-carboxymethyl inhibitor, His 231 and Tyr 157 are hydrogen bonded to the zinc-bound carboxylate, suggesting that these two groups might stabilize the tetrahedral transition state by donating hydrogen bonds in a manner analogous to the oxyanion stabilization in the serine proteases (Robertus et al., 1972). The observation that there are similar hydrogen bonds between oxygen PO1 of P-Leu-NH_2 and both His 231 and Tyr 157 provides additional support for this idea.

We do not see major conformational changes when P-Leu-NH_2 binds to thermolysin. Neither are there significant changes in the "disorder" or "thermal motion" of the residues in the active site when P-Leu-NH_2 or the N-carboxymethyl dipeptide inhibitor bind to thermolysin. The only adjustment in the protein that can be regarded as significant is a concerted "upwards" movement of about 0.25Å in the extended polypeptide segment that includes Ala 113 and Phe 114 (Fig. 17). This movement was noted previously with phosphoramidon (Weaver et al., 1977) and occurs for Phe-Leu-NH_2 as well. A similar but larger conformational shift was observed for a covalent modification of thermolysin (Holmes et al., 1983). The overall impression is that the active site is poised to accept an inhibitor or substrate, and undergoes relatively slight changes when such molecules are bound.

CHAPTER VI

SIMILAR PHOSPHONAMIDATE PEPTIDE INHIBITORS OF THERMOLYSIN DISPLAY DISSIMILAR MODES OF BINDING: A CRYSTALLOGRAPHIC ANALYSIS OF TWO PRESUMED TRANSITION STATE ANALOGS

Abstract

The modes of binding to thermolysin of two phosphonamidate peptide inhibitors, carbobenzoxy-Gly^{P}(OH)-L-Leu-L-Leu (ZG^{P}LL) and carbobenzoxy-L-Phe^{P}(OH)-L-Leu-L-Ala (ZP^{P}LA) have been determined by X-ray crystallography and refined at high resolution to crystallographic R-values of 17.7% and 17.0% respectively. [Gly^{P}(OH) is used to indicate that the trigonal carbon of the peptide linkage is replaced by the tetrahedral phosphorus of a phosphonamidate group.] These inhibitors were designed to be structural analogs of the presumed catalytic transition state and are potent inhibitors of thermolysin (ZG^{P}LL, K_i = 9.1 nM; ZP^{P}LA, K_i = 0.1 nM) (Bartlett & Marlowe, 1983; P.A. Bartlett, personal communication).

ZP^{P}LA is found to bind to thermolysin in the manner expected for the transition state and, for the first time, provides direct support for the presumed mode of binding of extended substrates in the S_2 subsite. The mode of binding of ZP^{P}LA displays all the interactions that are presumed to stabilize the transition state and supports the presumed mechanism of catalysis (Hangauer et al., 1984). The two

oxygens of the phosphonamidate moiety are liganded to the zinc to give overall pentacoordination of the metal.

For the second inhibitor the situation is different. Although both ZP^{P}LA and ZG^{P}LL have similar modes of binding in the S_1' and S_2' subsites, the configurations of the carbobenzoxy-Phe and carbobenzoxy-Gly moietys are different. For ZP^{P}LA the carbonyl group of the carbobenzoxy group is hydrogen bonded directly to the enzyme whereas in ZG^{P}LL the carbonyl group is rotated 117° and there is a water molecule interposed between the inhibitor and the enzyme. For ZG^{P}LL only one of the phosphonamidate oxygens is liganded to the zinc. Correlated with the change in inhibitor-zinc ligation from monodentate in ZG^{P}LL to bidentate in ZP^{P}LA there is an increase in the phosphorous-nitrogen bond length of about 0.25Å, strongly suggesting that the phosphonamide nitrogen in ZP^{P}LA is protonated, analogous to the protonated nitrogen of the transition state. The observation that the nitrogen of ZP^{P}LA appears to donate two hydrogen bonds to the protein also indicates that it is protonated. The different configurations adopted by the respective inhibitors are correlated with large differences in their kinetics of binding (P.A. Bartlett, personal communication). It is proposed that these differences in rates of binding can be rationalized in terms of the different configurational flexibility of the respective inhibitors and differences in bound solvent molecules that are displaced when the inhibitors bind to the enzyme.

Introduction

Thermolysin, isolated from the bacterium Bacillus thermoproteolyticus, is a zinc-requiring endopeptidase of molecular weight 34,600. The three-dimensional structure of the protein is known and has been refined to a nominal resolution of 1.6Å (Holmes & Matthews, 1982). Overall, the tertiary structure of thermolysin may be described as two spherical domains separated by a deep cleft that constitutes the active site.

It has become increasingly apparent that zinc-containing proteases are widely distributed in nature and play important roles in numerous physiological processes such as digestion and blood pressure regulation. In an attempt to understand the catalytic mechanism of thermolysin and other related zinc proteases, a series of inhibitors of the enzyme has been studied crystallographically (Kester & Matthews, 1977; Weaver et al., 1977; Bolognesi & Matthews, 1979; Holmes & Matthews, 1981; Monzingo & Matthews, 1982; Holmes et al., 1983; Monzingo & Matthews, 1984; Tronrud et al., 1986). Table 20 provides a summary of these various inhibitors and illustrates the manner in which they bind to thermolysin according to the "subsite" nomenclature introduced by Schechter and Berger (1967). Taken together, these inhibitor studies have suggested the catalytic mechanism shown in simplified form in Figure 21 (Hangauer et al., 1984). At the presumptive transition state the hydrated carbonyl carbon of the peptide substrate is in a tetrahedral configuration and the zinc ion is pentacoordinated by three protein plus

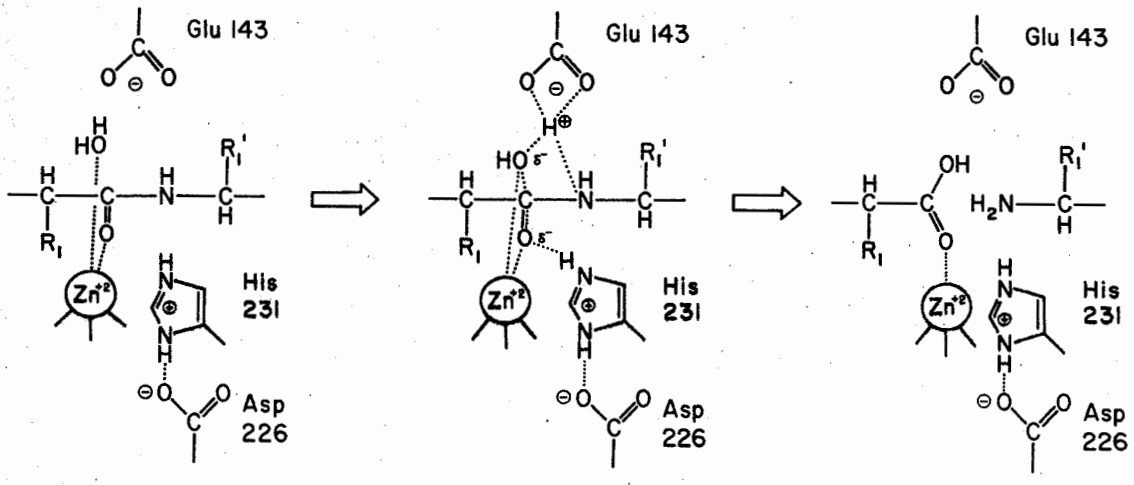


FIGURE 21. Schematic Representation of the Key Features of the Presumed Mechanism of Action of Thermolysin

An incoming peptide is presumed to displace a zinc-bound water molecule toward Glu 143, forming a pentacoordinate complex. The water molecule, activated by the combined influence of the metal ion and Glu 143, attacks the carbonyl carbon to form a tetrahedral intermediate. The tetrahedral intermediate is presumed to form a bidentate complex with the zinc and is stabilized in part by hydrogen bonds from His 231 and Tyr 157 (not shown in the figure). Glu 143 accepts the proton from the activated water molecule and is presumed to subsequently donate the proton to the leaving nitrogen, although proton donation by solvent water is not excluded as an alternative (after Hangauer et al., 1984).

two substrate ligands, rather than three protein ligands plus a water molecule as in native thermolysin.

The concept of transition state analogues has proven very effective as a basis for designing potent enzyme inhibitors. Such transition state analogues are synthesized on the premise that binding interactions between an enzyme and its substrate are optimal at the transition state (Pauling, 1946; Wolfenden, 1976; Bartlett & Marlowe, 1983). Apart from their potential as antimetabolites, transition state inhibitors can provide mechanistic information regarding enzyme catalysis (Wolfenden, 1976).

Phosphoramidon (N-(α-L-rhamnopyranosyl-oxyhydroxyphosphinyl)-L-Leu-L-Trp) is a potent, naturally-occurring, inhibitor of thermolysin (Suda et al., 1973; Komiyama et al., 1975) and is presumed to be a transition state analogue (Weaver & Matthews, 1977). Other, simpler, phosphoramides have also been shown to be very effective inhibitors of the zinc peptidases (Holmquist, 1977; Kam et al., 1979; Nishino & Powers, 1979; Holmquist & Vallee, 1979; Jacobsen & Bartlett, 1981; Galardy, 1980; Thorsett et al., 1982; Galardy et al., 1983). Extending these design principles, Bartlett and Marlowe have synthesized a series of phosphorous-containing peptide analogues designed specifically to mimic the geometry of an extended thermolysin substrate at the transition state (Bartlett & Marlowe, 1983; Bartlett, personal communication). In this report we describe the modes of binding to crystalline thermolysin of two of these inhibitors, namely carbobenzoxy-Gly^{P}(OH)-L-Leu-L-Leu ($ZG^{P}LL$) and carbobenzoxy-L-Phe^{P}(OH)-L-Leu-L-Ala ($ZP^{P}LA$) (Fig. 22). The

TABLE 20. Binding of Thermolysin Inhibitors Determined Crystallographically[a]

Inhibitor	K_i (μM)	S_2 R_2	S_1 R_1	Zinc ligand	S_1' R_1'	S_2' R_2'	S_3' R_3'
L-Benzylsuccinic acid	3800			$^-$OOC-	Benzyl		
L-Phenylalanyl-L-phenylalanyl amide	2000		Phe	-CO-	Phe		
β-Phenylpropionyl-L-phenylalanine	1600		β-Phenyl	-CO-	Phe		
Carbobenzoxy-L-phenylalanine	510		Phe	$^-$OOC-	Carbobenzoxy		
L-Leu-NHOH	190			HONCO-	Leu		
HONH-benzylmalonyl-L-Ala-Gly-p-nitroanilide	0.43			HONCO-	Benzyl	Ala	Gly
2-Benzyl-3-mercaptopropanoyl-L-alanylglycinamide	0.75		Benzyl	$^-$S-	Ala	Gly	
N-(1-Carboxy-3-phenylpropyl)-L-leucyl-L-tryptophan	0.05		Phenyl	$^-$OOC-	Leu	Trp	
Phosphoramidon	0.028		Rhamnose	-PO_2^--	Leu	Trp	
N-Phosphoryl-L-leucinamide	1.3			$^{2-}O_3P$-	Leu		
Cbz-Gly^P(OH)-L-Leu-L-Leu	0.0091	(Cbz)	(Gly)	-PO_2^--	Leu	Leu	
Cbz-L-Phe^P(OH)-L-Leu-L-Ala	0.0001	Cbz	Phe	-PO_2^-	Leu	Ala	

[a] The letters "S" and "R" refer to the active site subsites and the residues that occupy each subsite (after Schechter & Berger, 1967).

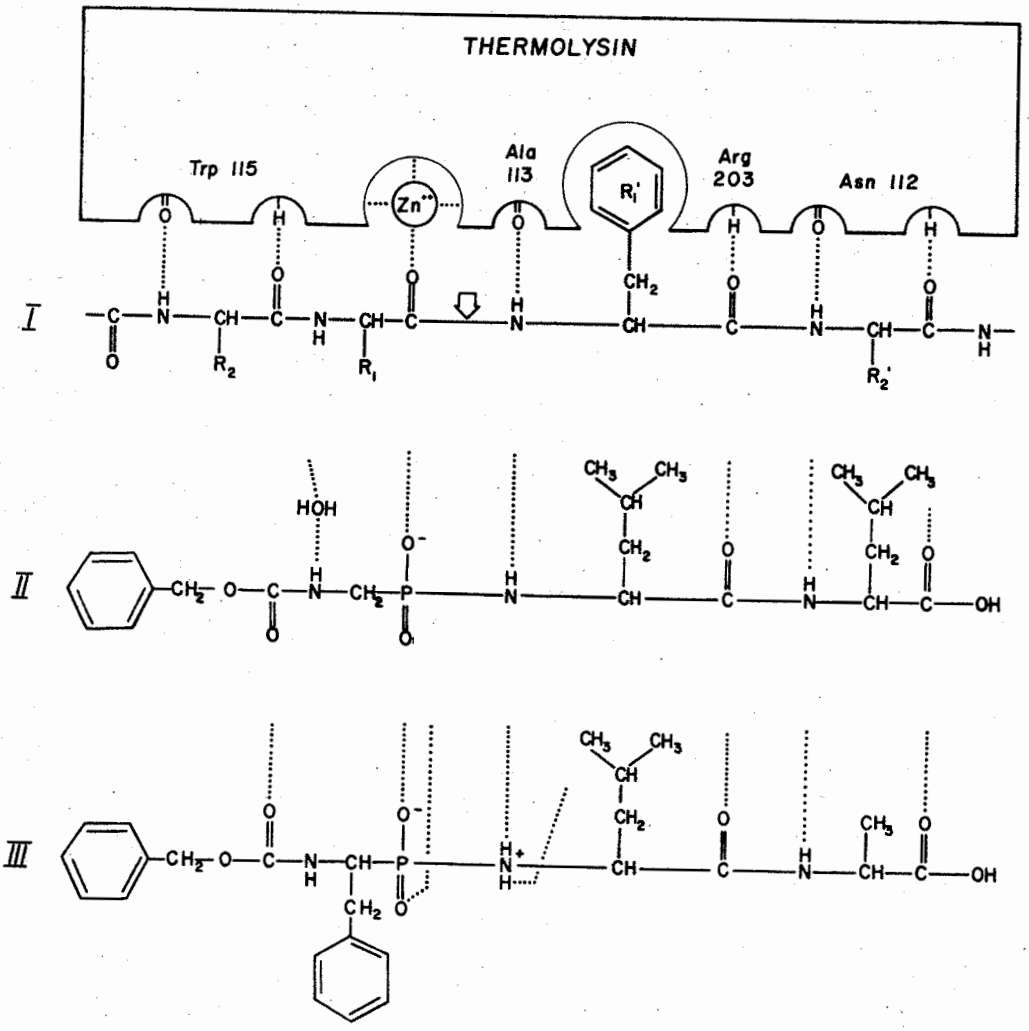


FIGURE 22. Schematic Illustration Showing the Structures of the Inhibitors ZG^{P}LL (II) and ZP^{P}LA (III) and Their Observed Interactions With the Extended Thermolysin Active Site

The figure includes an extended substrate (I) with its presumed hydrogen bonds to the enzyme shown as dotted lines and the bond to be cleaved indicated by the arrowhead. Note the different modes of interaction of the carbobenzoxy-Gly and the carbobenzoxy-Phe groups in II and III.

superscript "P" indicates the position where the tetrahedral phosphonamide moiety replaces the planar peptide linkage. The crystallographic results provide new data on the binding of extended substrates in the thermolysin active cleft and support the proposed mechanism of action (Monzingo & Matthews, 1984; Hangauer et al., 1984). In addition, the structural studies illustrate how inhibitors that have similar structures can adopt different modes of binding. In this instance, the alternative configurations are correlated with large differences in the rates of binding (Bartlett & Marlowe, 1983; P.A. Bartlett, personal communication).

Experimental Procedures

Thermolysin from Calbiochem was crystallized by the method of Holmes and Matthews (1982). The crystals grow as hexagonal rods and belong to the space group $P6_122$ with one monomer per asymmetric unit. They are stored in a synthetic mother liquor of 10 mM calcium acetate, 10 mM Tris, 7% (v/v) DMSO, pH = 7.2. The unit cell dimensions of native thermolysin crystals are $a = b = 94.2$Å, $c = 131.4$Å. The two inhibitors used in this study were gifts from Drs. C.K. Marlowe and P.A. Bartlett.

To prepare enzyme-inhibitor complexes, native thermolysin crystals were soaked at 4°C for a period of several days in either 10mM $ZG^{P}LL$ or 0.1mM $ZP^{P}LA$ dissolved in synthetic mother liquor. The lower concentration of $ZP^{P}LA$ was used because of the tendency of the crystals to crack when exposed to this inhibitor. Even at the lower concentration the crystals still develop some cracks. Inhibitor binding was monitored by

calculating (hOl) difference Fourier projection maps based on the appropriate precession photographs (Weaver et al., 1977).

A three-dimensional data set was collected for each enzyme-inhibitor complex by the method of oscillation photography (Schmid et al., 1981). The X-ray source was a graphite-monochromatized Elliot GX-21 rotating anode generator operated at 39 kV and 120 mA. An oscillation angle of 1.2° per film pack was used and the crystal was rotated about the c-axis through a net rotation of 30°. The typical exposure time was 4 or 6 hours per film pack. By making an appropriate translation, only one crystal was required for each data set. Data to 1.6Å resolution were measured for the thermolysin:$ZG^{P}LL$ complex and to 1.7Å resolution for the thermolysin:$ZP^{P}LA$ complex. Data collection statistics are summarized in Table 21.

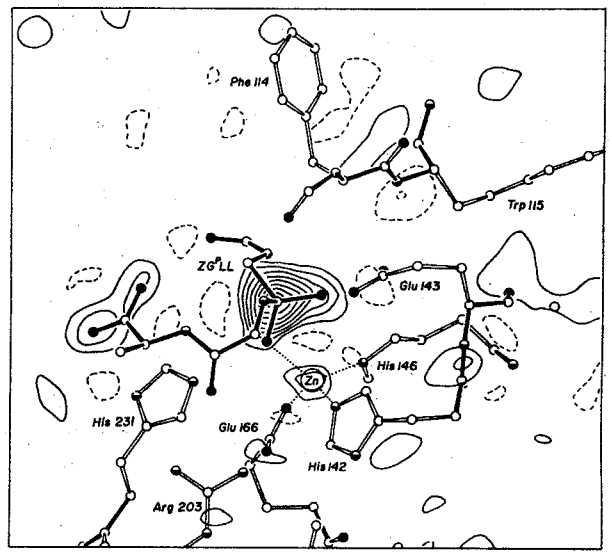
Difference electron density maps with amplitudes of the form ($F_{complex} - F_{nat.,calc}$) and phases calculated from the refined native structure clearly showed the respective inhibitors binding in the thermolysin active site cleft with the electron-dense phosphorous atom adjacent to the zinc (Fig. 23). The height of the peak corresponding to the phosphorus was 17σ for $ZG^{P}LL$ and 10σ for $ZP^{P}LA$ where σ is the root mean square value of the difference density throughout the unit cell. [Difference density maps with coefficients of the form ($F_{complex} - F_{nat.,obs}$) also show inhibitor binding but their interpretation is confused by the displacement of solvent molecules (cf. Kester & Matthews, 1977).]

Initial coordinates for each inhibitor were obtained by modeling the inhibitor into the electron density on an Evans and Sutherland PS330

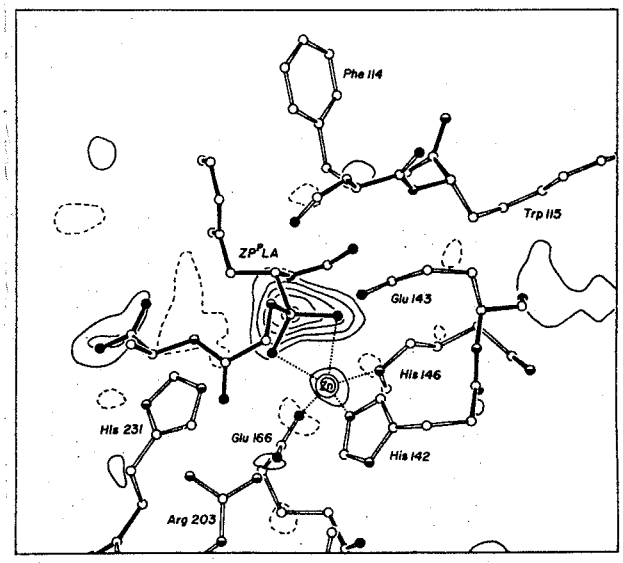
TABLE 21. Intensity Statistics for $ZG^{P}LL$ and $ZP^{P}LA$

	$ZG^{P}LL$	$ZP^{P}LA$
Number of films	28	27
Average $R_{sym}^{(a)}$ (%)	3.3	2.7
Average $R_{sca}^{(a)}$ (%)	6.4	4.8
$R_{merge}^{(a)}$ (%)	5.6	4.2
Total reflections measured	72,525	58,214
Independent reflections	31,799	28,875
Resolution (Å)	1.6	1.7
Average isomorphous difference (%)	16.5	11.5
Cell dimensions		
$\underline{a},\underline{b}$ (Å)	94.1	93.8
$\underline{c}$ (Å)	132.0	131.9

(a) $R = \Sigma \left| I-\overline{I} \right| \Sigma I$. R_{sym} measures the agreement between symmetry-related reflections on the same film, R_{sca} measures the agreement between reflections recorded on successive films in a given film pack and R_{merge} gives the overall agreement between intensities measured on different films.



(a)



(b)

FIGURE 23. Section of Difference Map Cutting Through the Bound Inhibitor ZG^{P}LL and ZP^{P}LA

Section z = -0.044 of a 1.9Å resolution difference electron density maps with amplitudes ($F_{Complex} - F_{Nat.,calc.}$) and phases calculated from the refined native structure. Contours are drawn at intervals of 2.0σ with positive contours solid and negative contours broken. Protein and inhibitor atoms lying within ±3.5Å of the section are included; oxygen atoms are drawn solid, nitrogen half-solid and carbon as open circles. Inhibitor bonds and protein backbone bonds are drawn solid. (a) ZG^{P}LL. (b) ZP^{P}LA.

graphics system. Electron density maps used for the initial model-building studies were based on ($2F_{complex} - F_{nat.,calc}$) coefficients. Starting coordinates for refinement were these crude coordinates plus the coordinates of native thermolysin refined to 1.6Å resolution (Holmes & Matthews, 1982). Refinement of the thermolysin:inhibitor complexes was carried out using the "TNT" system of programs written and developed in this laboratory (see Chapter III). These programs are based on the method of restrained least squares. Occasionally during the course of the refinement, electron density maps based on ($F_{complex} - F_{calc}$) coefficients were checked for possible water molecules missing from the initial protein:inhibitor coordinates. "Ideal" stereochemistry for the protein and the inhibitor was based on the values in the "TNT" refinement package as well as bond lengths and angles from Bowen et al. (1958) and Kojima et al. (1978). Results from the refinement are presented in Tables 22-24 and the refined coordinates for the thermolysin-inhibitor complexes have been deposited in the Brookhaven Data Bank.

Results

Binding of Inhibitors to Crystalline Thermolysin

The observed binding of $ZG^{P}LL$ in the thermolysin active site is shown in Figure 24. As expected, the inhibitor binds in an extended conformation with the two leucine residues occupying the S_1' and S_2' specificity pockets. The zinc:phosphate-oxygen distances are 3.0Å and 2.1Å. Additional details of the geometry of the zinc ligands are given in Table 25. Relevant protein:inhibitor contacts are given in Table 26.

TABLE 22. Refinement Statistics for $ZG^{P}LL$ and $ZP^{P}LA$

	$ZG^{P}LL$	$ZP^{P}LA$
Resolution limits (Å)	10.0 - 1.6	10.0 - 1.7
Initial R-factor (%)[a]	24.9	21.6
Final R-factor (%)	17.7	17.0
Number of cycles	28	24
Number of reflections used	31,627	28,691
Number of atoms	2,643	2,635
Weighted rms deviations from ideality		
Bond length (Å)	0.022	0.023
Bond angle (deg)	2.8	3.2
Planarity (trigonal) (Å)	0.010	0.012
Planarity (other planes) (Å)	0.014	0.015
Torsion angle (deg)	16.4[b]	16.5[b]

[a] R-factor = $\Sigma \mid F_{obs} - F_{calc} \mid / \Sigma \mid F_{obs} \mid$

[b] The torsion angles were not restrained during refinement.

TABLE 23. Coordinates for ZG^{P}_{LL}

Atom	X	Y	Z	B
Carbobenzoxy				
OA	49.4	18.8	-9.9	28.1
CB	48.9	18.3	-11.2	29.9
CG	47.5	19.0	-11.6	25.9
CD1	46.5	19.0	-10.7	26.5
CE1	45.3	19.6	-11.0	25.0
CZ	45.1	20.3	-12.1	23.4
CE2	46.2	20.4	-13.0	26.7
CD2	47.4	19.8	-12.7	29.5
C	49.7	17.9	-8.9	20.9
O	49.6	16.7	-9.0	20.4
Glycine				
N	50.1	18.6	-7.9	10.7
CA	50.4	18.0	-6.7	3.8
Phosphonamide				
P	51.7	19.0	-6.1	9.7
OP1	53.1	18.6	-6.6	9.8
OP2	51.5	20.5	-6.1	10.6
N	51.7	18.5	-4.5	2.6
Leucine				
CA	52.8	18.2	-3.6	13.4
CB	52.5	18.3	-2.1	5.6
CG	53.7	18.2	-1.2	11.5
CD1	54.6	19.4	-1.4	10.5
CD2	53.2	18.1	0.2	7.7
C	53.5	16.9	-3.9	19.9
O	54.8	16.8	-3.8	9.2
N	52.7	15.9	-4.2	11.7
Leucine				
CA	53.2	14.5	-4.5	20.3
CB	53.5	13.7	-3.1	11.7
CG	52.3	13.5	-2.3	17.8
CD1	52.3	14.5	-1.2	23.6
CD2	52.3	12.2	-1.5	24.7
C	52.4	13.8	-5.5	22.0
O	51.3	14.5	-5.9	11.4
OH	52.8	12.8	-6.1	18.8

TABLE 24. Coordinates for $ZP^{P}LA$ and Discrepancy with $ZG^{P}LL$

Atom	X	Y	Z	Δr	B	ΔB
Carbobenzoxy						
OA	50.8	20.6	-9.9	2.3	36.8	9
CB	51.0	21.9	-10.4	4.2	28.9	-1
CG	50.5	21.8	-11.8	4.0	35.8	10
CD1	50.4	20.6	-12.6	4.6	40.1	14
CE1	49.9	20.6	-13.9	5.5	35.9	11
CZ	49.5	21.7	-14.5	5.1	34.7	11
CE2	49.5	22.8	-13.7	4.1	36.7	10
CD2	50.0	22.9	-12.4	4.1	45.5	16
C	50.4	20.4	-8.6	2.6	34.3	13
O	49.9	21.3	-7.8	4.7	14.8	-6
Phenylalanine						
CB	50.7	17.2	-7.0		13.7	
CG	49.4	16.8	-7.7		21.8	
CD1	48.2	16.8	-6.9		18.6	
CE1	47.0	16.5	-7.5		29.1	
CD2	49.3	16.6	-9.0		24.6	
CE2	48.1	16.3	-9.6		28.5	
CZ	46.9	16.2	-8.8		30.9	
N	50.8	19.1	-8.2	0.9	22.5	12
CA	50.7	18.6	-6.8	0.7	15.6	12
Phosphoramide						
P	52.0	19.2	-6.1	0.4	21.7	12
OP1	53.3	18.5	-6.7	0.3	16.8	7
OP2	52.1	20.7	-5.9	0.7	20.5	10
N	51.7	18.5	-4.3	0.2	10.7	8
Leucine						
CA	52.9	18.3	-3.5	0.2	19.6	6
CB	52.6	18.5	-2.0	0.2	11.4	6
CG	53.8	18.4	-1.1	0.2	18.7	7
CD1	54.7	19.7	-1.3	0.4	11.0	0
CD2	53.2	18.2	0.3	0.1	14.5	7
C	53.6	17.0	-3.6	0.3	21.3	1
O	54.9	16.9	-3.6	0.2	13.7	4
N	52.9	15.9	-3.8	0.4	15.4	4

TABLE 24. (continued)

Alanine

CA	53.3	14.5	-3.9	0.6	19.5	-1
CB	53.0	13.8	-2.6	0.7	14.2	3
C	52.7	13.8	-5.1	0.6	27.6	6
O	51.7	14.3	-5.6	0.4	23.9	13
OH	53.1	12.8	-5.7	0.5	37.9	19

Coordinates are in Angstroms in the standard orthogonal thermolysin coordinate system (Matthews et al., 1974). B is the crystallographic thermal parameter in $Å^2$. Δr and ΔB are the differences between the atomic positions and the thermal factors for ZP^{P}LA and ZG^{P}LL.

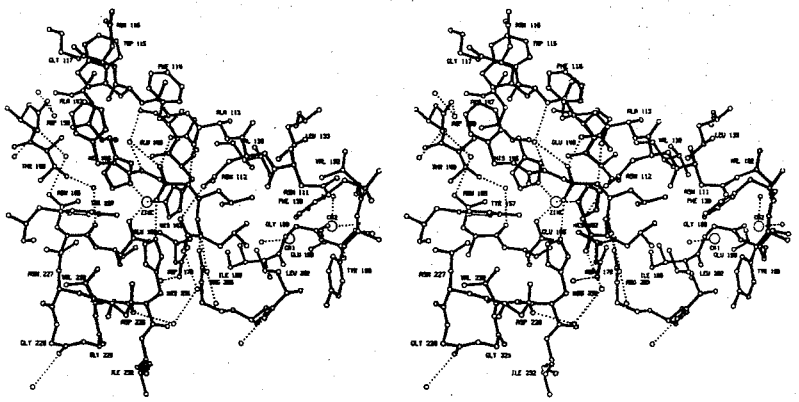


FIGURE 24. Stereo Drawing Showing the Mode of Binding of ZG^{P}LL in the Extended Thermolysin Active Site Cleft

The inhibitor and the main-chain bonds of the protein are drawn solid; protein side chains have open bonds. Apparent hydrogen bonds between the inhibitor and the protein are shown as dotted lines. A number of bound water molecules are also included.

Figure 25 shows stereographically the binding of $ZP^{P}LA$ to thermolysin. Again, relevant protein:inhibitor contacts are given in Tables 25 and 26. Like $ZG^{P}LL$, $ZP^{P}LA$ binds to the enzyme with its phosphate oxygens close to the zinc and the leucine and alanine residues occupying the S_1' and S_2' subsites respectively. However, in the case of $ZP^{P}LA$, the coordination of the zinc by the phosphate oxygens is decidedly bidentate with zinc:phosphate distances of 2.2Å and 2.6Å.

As can be seen in Figure 26, the R_1' and R_2' residues of both inhibitors bind to thermolysin in nearly the same orientation and replacement of leucine by alanine in $ZP^{P}LA$ makes little difference to the mode of binding in the S_1' and S_2' subsites. However, as can also be seen from Figure 26, and is shown schematically in Figure 22, $ZG^{P}LL$ and $ZP^{P}LA$ bind to thermolysin in a completely different orientation with respect to their carbobenzoxy moieties. In fact, the dihedral angle defined by the N-C_α bond of inhibitor residue R_1 differs between the two analogues by 117°.

Discussion

Mode of Binding of Extended Substrates

Previous crystallographic analyses of the binding of peptide-analogue inhibitors to thermolysin have illustrated how extended substrates bind in the S_1, S_1' and S_2' subsites (e.g. see Weaver *et al.*, 1977; Holmes & Matthews, 1981; Monzingo & Matthews, 1982, 1984). In a model first proposed by Kester and Matthews (1977) and elaborated by Hangauer *et al.* (1984) it was suggested that in the S_1-S_2 subsites, the

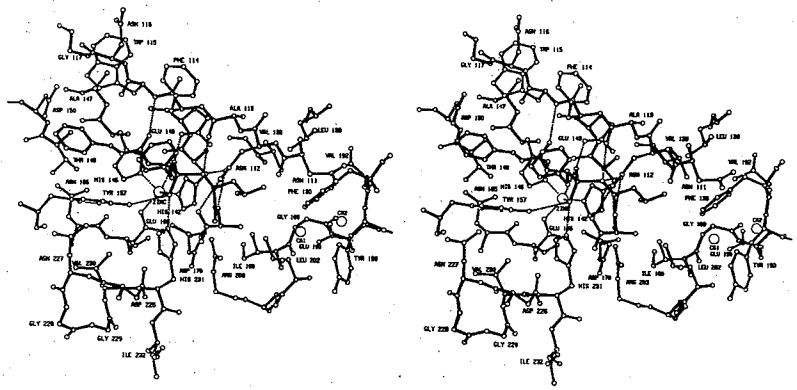


FIGURE 25. The Binding of ZP^{P}LA to Thermolysin.

TABLE 25. Zinc-Ligand Geometry in ZG^{P}LL and ZP^{P}LA

Ligand	Distance in ZG^{P}LL Complex (Å)	Distance in ZP^{P}LA Complex (Å)
His 142 NE2	2.04	2.09
His 146 NE2	2.09	2.11
Glu 166 OE1	2.07	2.04
Inhibitor OP1	2.08	2.17
Inhibitor OP2	3.01	2.59

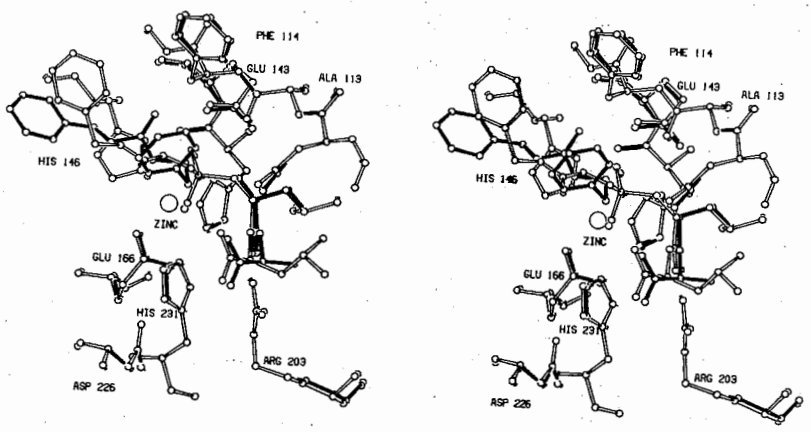


FIGURE 26. Comparison of Binding of ZG^{P}LL (Open Bonds) and ZP^{P}LA (Solid Bonds)

polypeptide backbone of the substrate hydrogen bonded with the protein backbone of Trp 115 as in an antiparallel β-sheet (Fig. 22). The binding of $ZP^{P}LA$ provides the first experimental support for this proposed mode of interaction. In particular, the carbonyl oxygen of the inhibitor carbobenzoxyl moiety makes a good hydrogen bond (2.9Å) with the peptide nitrogen of Trp 115 (Fig. 25). In order to achieve this mode of binding a water molecule (Sol 362), which normally hydrogen bonds to Trp 115 in the native enzyme, is expelled from the active site.

In the case of $ZG^{P}LL$, however, the interactions with the enzyme in the S_2 subsite are very different. In this case the carbonyl oxygen of the inhibitor points away from the amide group of Trp 115. The water molecule, Sol 362, is not displaced and forms a hydrogen bonded bridge between the amide nitrogen of the inhibitor R_1 residue and the amide of Trp 115.

Mechanism of Catalysis

A mechanism for the cleavage of peptides by thermolysin was first proposed by Pangburn and Walsh (1975) and by Kester and Matthews (1977) and has been elaborated by Holmes and Matthews (1981), Monzingo and Matthews (1984) and Hangauer et al. (1984) (see also Antonov et al., 1981; Kunugi et al., 1982). According to the mechanism based on these studies and outlined in Figure 21, the carbonyl oxygen of the incoming substrate coordinates to the zinc ion of the protein. The water molecule, normally the fourth ligand to the zinc in native thermolysin, is displaced toward Glu 143 but is not totally excluded from the zinc

coordination sphere. There is then a nucleophilic attack by the water molecule on the carbonyl carbon of the substrate. The attacking water molecule is thought to be activated by interactions with both Glu 143 and the zinc ion. Recent studies (Monzingo & Matthews, 1984; Hangauer et al., 1984) suggest that Glu 143 accepts a proton from the attacking water molecule and shuttles this proton to the leaving nitrogen. In the presumed transition state the carbonyl carbon of the substrate is tetrahedral and the zinc ion is coordinated to five ligands rather than four as in the ground state. As discussed below, the observed mode of binding of $ZP^{P}LA$ to thermolysin provides support for several aspects of the proposed mechanism.

$ZP^{P}LA$ is the tightest inhibitor of thermolysin described to date (K_i = 0.1 nM; Bartlett, personal communication) and its interactions with the enzyme are presumed to resemble those that occur for a true substrate in the transition state. The interactions of the phosphonamide group are of particular interest. One of the phosphonamide oxygens accepts hydrogen bonds from both Tyr 157 and His 231 (Table 26). Such hydrogen bonds could stabilize the tetrahedral (hydrated carbon) intermediate during catalysis, analogous to the "oxyanion hole" in the serine proteases (cf. Hangauer et al., 1984; see also Chapter V). The second phosphonamide oxygen is very close (2.3Å) to one of the carboxyl oxygens of Glu 143 (Table 26). (The refinement program pushes apart pairs of oxygen atoms only if their interatomic distance is less than 2.3Å.) We interpret the close approach distance as indicating that the carboxyl group of Glu 143 is protonated and donates a strong hydrogen

TABLE 26. Selected Thermolysin-Inhibitor Distances for $ZG^{P}LL$ and $ZP^{P}LA$[a]

Protein	Inhibitor		Distance (Å)	
	$ZG^{P}LL$	$ZP^{P}LA$	$ZG^{P}LL$	$ZP^{P}LA$
Trp 115 NH	Sol 362	Cbz O	2.8(H)	3.0(H)
Tyr 157 OH	Gly NH	Gly NH	3.9	3.3
Tyr 157 OH	OP1	OP1	3.4	3.1(H)
His 231 NE2	OP1	OP1	2.9(H)	2.7(H)
Glu 143 OE1	OP2	OP2	2.5(H)	2.3(H)
Glu 143 OE2	Leu(1) N	Leu N	3.4	3.3
Ala 113 O	Leu(1) N	Leu N	3.0(H)	2.9(H)
Asn 112 OD1	Leu(1) N	Leu N	3.3	3.0(H)
Arg 203 NH1	Leu(1) O	Leu O	3.0(H)	2.8(H)
Arg 203 NH2	Leu(1) O	Leu O	3.0(H)	2.9(H)
Asn 112 OD1	Leu(2) N	Ala N	3.2(H)	3.1(H)
Asn 112 ND2	Leu(2) O	Ala O	3.0(H)	3.0(H)

(a)(H) indicates a presumed hydrogen bond.

bond to the phosphonamide oxygen. [Carboxyl oxygens are known to form strong hydrogen bonds with oxygen-oxygen distances as short as 2.4-2.5Å (Jeffrey & Maluszynska, 1982).] The same type of hydrogen bonding would be expected to occur between Glu 143 and the transition state intermediate formed by attack of the activated water molecule on the carbonyl carbon (Fig. 21). Because the hydrated peptide of the transition state is less bulky than the phosphonamide group, the hydrogen bond between Glu 143 and the transition state could be longer than the value of 2.3Å observed for ZP^{P}LA.

It is presumed that in the transition state the zinc ion is pentacoordinate. This was first suggested by the observation that hydroxamic acid inhibitors show five-fold coordination (Holmes & Matthews, 1981) and is also supported by the bidentate ligation of N-carboxymethyl peptide inhibitors (Monzingo & Matthews, 1984). In the present study, ZP^{P}LA, the tighter-binding inhibitor and the presumed transition state analogue, also shows five-fold coordination, whereas ZG^{P}LL, which is presumed to bind in a "non-productive" mode, has essentially tetrahedral coordination at the metal. Tetrahedral coordination was also observed for the binding of phosphoramidon (Weaver & Matthews, 1977; see also Chapter V). Zinc-ligand distances for a series of thermolysin-inhibitor complexes are summarized in Table 27. We presume that as the zinc ligation comes closer to full pentacoordination, the phosphorous-containing inhibitors become better approximations to the transition state. The phosphonamide groups of phosphoramidon and ZP^{P}LA are shown superimposed in Figure 27. These are the extreme examples

TABLE 27. Geometry of Different Phosphonamide Inhibitors of Thermolysin

Inhibitor	Zinc-ligand distance (A)		Liganding asymmetry	P-N bond length (A)	P-N-C angle	Nitrogen protein distances (A)			Reference
	OP1	OP2	$\frac{OP2-OP1}{OP2+OP1}$			Asn 112	Ala 113	Glu 143	
Phosphoramidon	1.75	3.40	0.32	1.42	128°	3.4	3.4	3.9	Tronrud et al. (1986)
ZGPLL (1)	2.08	3.01	0.18	1.66	129°	3.3	3.0	3.4	This work
ZGPLL (2)[a]	(1.90)	(3.17)	(0.22)	(1.71)	(121°)	(3.2)	(3.0)	(3.5)	This work
P-Leu-NH_2 (1)	2.06	2.79	0.15	1.85	122°	3.1	3.0	3.4	Tronrud et al. (1986)
P-Leu-NH_2 (2)[a]	(1.96)	(2.83)	(0.18)	(1.84)	(126°)	(3.7)[b]	(3.0)	(3.2)	Tronrud et al. (1986)
ZPPLA	2.17	2.59	0.09	1.90	115°	3.0	3.0	3.3	This work

(a) The repeated entries in the table were obtained from independent refinements of these inhibitors by two different procedures. In the first case the thermolysin-inhibitor complex was refined by "TNT", a restrained least-squares procedure (D.E.T., L.F. Ten Eyck and B.W.M., manuscript in preparation). The alternative refinement was by "EREF", a combined energy minimization-crystallographic refinement procedure (Jack & Levitt, 1971). These refinement programs are different in their overall philosophy. Also their dictionaries of "standard geometry" are different. In addition the intensity data were remeasured between the two refinements and the resolution was extended from 1.9Å for EREF to 1.7Å or 1.6Å for TNT. A comparison of the repeated observations in Table 27 gives an impression of the uncertainty due to the limitations of the refinement.

(b) The side chain of Asn 112 is in a slightly different conformation in the "EREF models". We believe the "TNT model" has the correct conformation for this side chain.

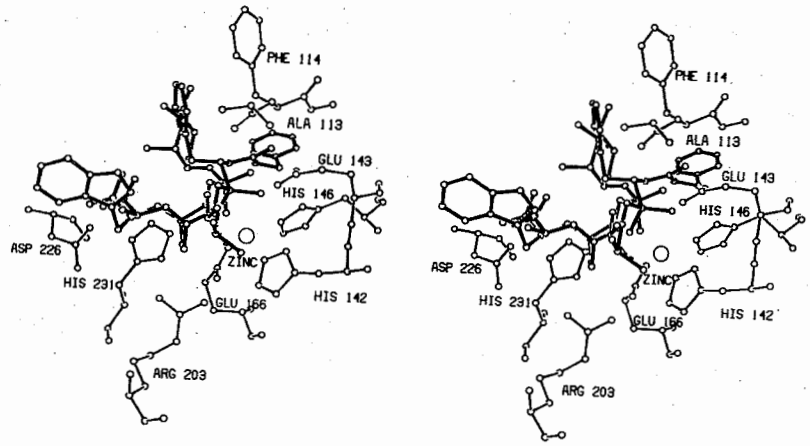


FIGURE 27. Comparison of Binding of Phosphoramidon (Open Bonds) and ZP^{P}LA (Solid Bonds)

from Table 27. $ZG^{P}LL$ is reasonably similar to phosphoramidon while N-phosphoryl-L-leucinamide (P-Leu-NH_2) is intermediate between $ZG^{P}LL$ and $ZP^{P}LA$ (Table 27).

The change in the coordination of the phosphonamide oxygens at the zinc corresponds, in part, to a rotation about the phosphorous-nitrogen bond (Fig. 27). This rotation is correlated with an apparent increase in the phosphorous-nitrogen bond length (Table 27), already noted in the comparison of phosphoramidon with P-Leu-NH_2 (Chapter V). (In all the crystallographic refinements the P-N bond was restrained to an "ideal" value of 1.78Å but, because of the uncertainty of the "expected" value for the bond length, was given a weight of 1/9 relative to all other bonds.) The increase in the P-N bond length (Table 27) can be taken as an indication that the nitrogen is becoming protonated. This is also suggested by the decrease in the angle at the nitrogen (from 129° in $ZG^{P}LL$ to 115° in $ZP^{P}LA$). If, indeed, the phosphonamide nitrogen in $ZP^{P}LA$ is protonated, as seems to be the case, it is yet another attribute of the bound inhibitor as an excellent mimic of the transition state.

Table 27 also includes, for the different inhibitors, the distances between the phosphonamide nitrogen and the nearest protein atoms. These distances (and the corresponding directions) show that there are oxygens available (Asn 112 OD1 and the peptide oxygen of Ala 113) to accept <u>two</u> hydrogen bonds from the nitrogen. For $ZP^{P}LA$ and P-Leu-NH_2, in which the nitrogen is presumed to be protonated, both these hydrogen bonds are realized. It will also be noted in Table 27 that as one proceeds from phosphoramidon to P-Leu-NH_2 and $ZP^{P}LA$, the nitrogen occupies a position

closer to Glu 143, consistent with the proposal that Glu 143 is the proton donor and subsequently stabilizes the protonated transition site by an ionic interaction. Because the two protein oxygens OD1 of Asn 112 and the peptide oxygen of Ala 113, already occupy the requisite positions to accept hydrogen bonds from phosphoramide nitrogen, it follows that these positions are not accessible to solvent without a conformational adjustment. Indeed, the mode of binding adopted by ZP^{P}LA (Fig. 25) is such that any access of water to the nitrogen is sterically excluded. Assuming that a substrate adopts a similar configuration, solvent would be excluded from the scissile nitrogen in the same way (also see Hangauer et al., 1984). This inaccessibility to solvent argues against water as the proton donor in catalysis, although it is difficult to rule out structural changes that would permit access to solvent.

As mentioned, the distance in ZP^{P}LA between the phosphonamide nitrogen and one of the oxygens of Glu 143 is 3.3Å (Table 26) consistent with the mechanistic proposal that Glu 143 accepts a proton from the attacking water molecule and subsequently shuttles the proton to the leaving nitrogen of the scissile peptide (Monzingo & Matthews, 1984; Hangauer et al., 1984). By analogy with thermolysin it was proposed that a similar mechanism of action should be considered for carboxypeptidase A (Monzingo & Matthews, 1984). In particular, it was suggested that in carboxypeptidase A the role of the proton donor could be ascribed to Glu 270 rather than Tyr 248. The recent demonstration that Tyr 248 of carboxypeptidase A is not essential for catalysis (Gardell et al., 1985) is consistent with the above suggestion and also

supports the idea that the mechanisms of action of carboxypeptidase A and thermolysin could be very similar (Kester & Matthews, 1977; Monzingo & Matthews, 1984).

Based on the observed mode of binding to thermolysin of N-carboxymethyl-dipeptide and other inhibitors, Hangauer et al. (1984) used interactive computer graphics to model the presumed intermediates that occur during catalysis. In Figure 28 we have superimposed that proposed tetrahedral transition state on the observed coordinates for $ZP^{P}LA$. As can be seen, the overall agreement is only moderately good and there are substantial differences in the binding of the leucine within the S_1' "specificity pocket". [The same discrepancy occurred between the proposed transition state model and the observed binding of "CLT" in the specificity pocket (see Fig. 7 of Hangauer et al., 1984).]

Kinetics of Binding

In general, transition state inhibitors bind more tightly to an enzyme than ground state analogues and often display an apparent slow rate of binding (Wolfenden, 1976; Frieden et al., 1980). Both $ZG^{P}LL$ and $ZP^{P}LA$ were designed specifically as transition state analogue inhibitors of thermolysin. Not surprisingly, these two inhibitors have different binding constants, but, unexpectedly, they differ with respect to their binding kinetics (Bartlett & Marlowe, 1983; Bartlett, personal communication). $ZG^{P}LL$ has an inhibition constant, K_i, of 9.1 nM and shows instantaneous binding kinetics. On the other hand $ZP^{P}LA$, the most potent inhibitor yet reported for thermolysin (Table 20) demonstrates

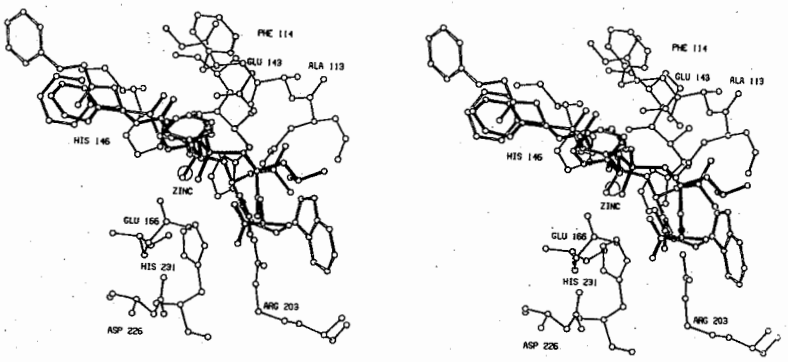


FIGURE 28. Comparison of Binding of ZP^{P}LA (Solid Bonds) and a Model for the Transition State (Hangauer et al., 1984)

slow binding kinetics with a K_i of 0.1 nM. The different properties of ZG^PLL and ZP^PLA observed in solution can now be correlated with the different modes of binding observed crystallographically.

The analogue ZP^PLA is presumed to represent the transition state both in its bidentate coordination at the zinc and its interactions with the enzyme in the S_2, S_1, S_1', and S_2' subsites. ZG^PLL is presumed to bind in a non-productive mode (at least in the S_1-S_2 region). The first question is why the two inhibitors adopt different modes of binding. The presence of the phenylanalyl side chain at the R_1 position has two consequences. First, it can make more extensive interactions with the enzyme than can a glycine and, second, it restricts those configurations accessible to the inhibitor much more than is the case with a glycine at this position. The possible configurations can be evaluated in exactly the same way as for a polypeptide (Ramachandran et al., 1963) except that the trigonal carbonyl group is replaced by the tetrahedral phosphonamide group. The results of rotating about the N-C_α bond of residue R_1 are shown in Figure 29. If R_1 is a non-glycine, the dihedral angle ϕ is relatively restricted, with allowed values from about -170° to -80°. The observed value for ZP^PLA ($R_1 \equiv$ Phe) is -83°. As expected, if R_1 is a glycine, much greater rotational flexibility is allowed (from about 100° through ±180°). The observed dihedral value for ZG^PLL is ϕ = 146°. Since the glycine-containing inhibitor could, in principle, adopt the conformation of its phenylalanyl counterpart, the differences in the modes of binding of the two inhibitors cannot be attributed to differences in their allowed conformations. Rather, the difference between the binding of ZP^PLA and ZG^PLL must be attributed to the additional

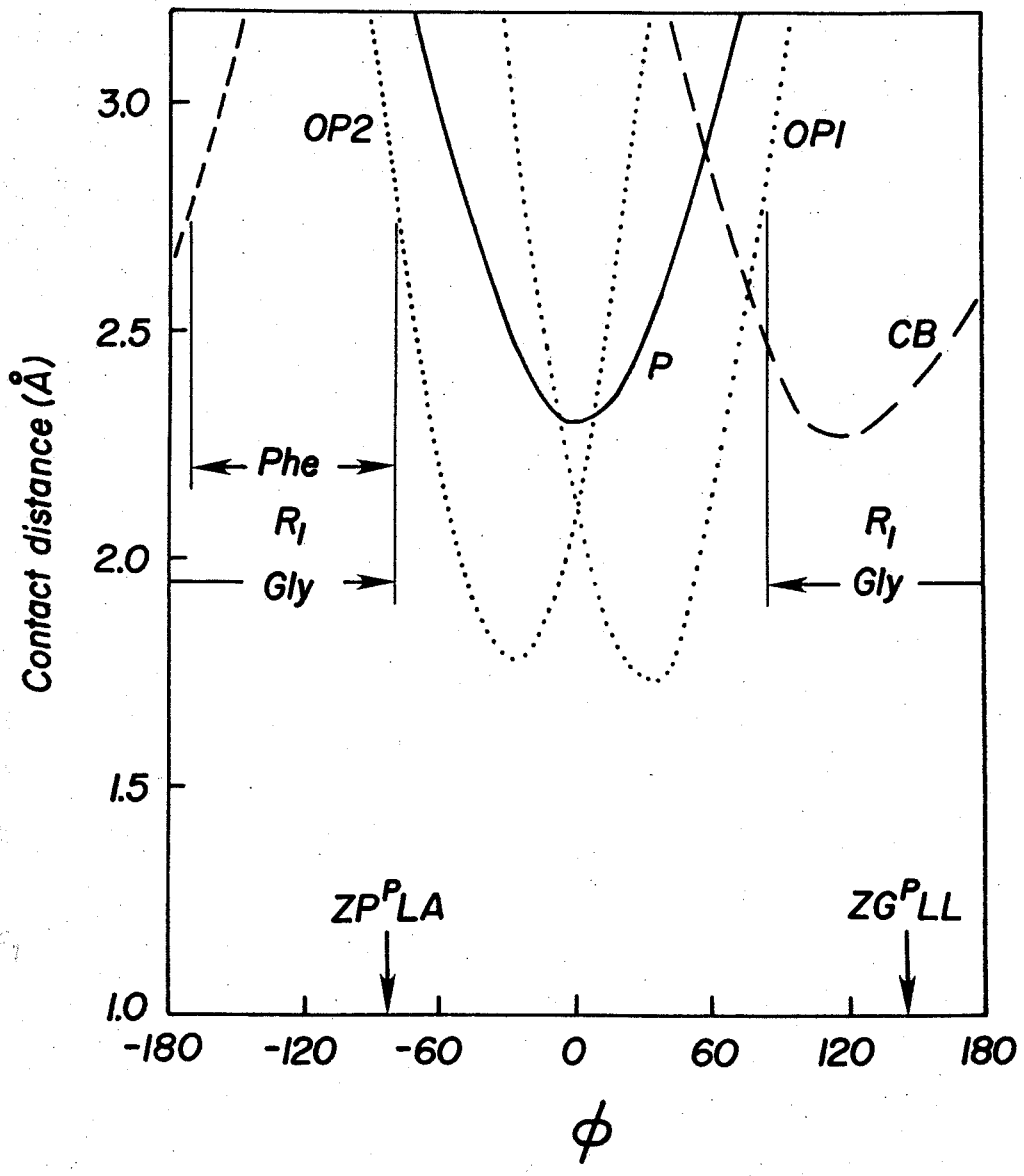


FIGURE 29. Selected Interatomic Distances in ZP^{P}LA as a Function of Rotation About the N-C_α (R1) Bond

Interatomic distances in ZP^{P}LA between the carbonyl oxygen of the carbobenzoxy group and, respectively, the phosphorous atom P (solid line), the phosphonamide oxygens OP1 and OP2 (dotted lines) and the β-carbon, CB, of the phenylalanyl side chain (dashed line). Each distance is plotted as a function of the rotation, ϕ, of the carbonyl group about the N-C_α (R_1) bond. Assuming that the shortest permissible interatomic approach is 2.7Å, the allowed ranges of values of ϕ for a glycine and a non-glycine (e.g. phenylalanine) at R_1 are as shown in the figure. The observed values for ZG^{P}LL and ZP^{P}LA are also shown.

interactions generated by the phenylalanyl side chain. Morihara and Tsuzuki (1970) showed that the K_M for the thermolysin substrate Z-Gly-Leu-Ala is 10.6 mM whereas Z-Phe-Leu-Ala has a K_M of 0.6 mM. Thus the phenylalanine side chain at the R_1 position clearly contributes to binding. In terms of the binding observed for ZP^{P}LA (Fig. 25) the phenylalanyl side chain lies at an angle to the face of Phe 114, an interaction also seen with other thermolysin inhibitors (e.g. see Kester & Matthews, 1977; Monzingo & Matthews, 1984). This additional interaction must account for the tighter binding of ZP^{P}LA relative to ZG^{P}LL, and, in addition, the different mode of binding. [Because the configuration of ZP^{P}LA in solution is more constrained than that of ZG^{P}LL (see below) there is also an entropic contribution to the tighter binding of ZP^{P}LA relative to ZG^{P}LL.] It might also be noted that the substitution of Ala for Leu in the R_2' position is expected to slightly weaken the binding of ZP^{P}LA relative to ZG^{P}LL. Morihara and Tsuzuki (1970) found Z-Gly-Leu-Leu to have a K_M of 2.6 mM whereas Z-Gly-Leu-Ala has a four-fold larger value (10.8 mM).

With respect to the difference in the kinetics of binding of the two inhibitors, we note that the binding of ZP^{P}LA requires the displacement of a number of water molecules, in particular the one hydrogen bonded to the peptide nitrogen of Trp 115 (i.e. in the S_2 subsite) (Fig. 25). In contrast, this solvent molecule is not displaced when ZG^{P}LL is bound (Fig. 24). Furthermore, as we have shown, the part of ZP^{P}LA that occupies the S_1 and S_2 subsites is much less flexible than is the case for ZG^{P}LL. Therefore the bound solvent in the S_2 subsite may have to be displaced well away from its ground-state position before ZP^{P}LA can

enter the active site. But this, in turn, would require a large activation energy, explaining the slow kinetics of binding for this inhibitor. ZG^{P}LL, on the other hand, being more flexible, is more free to change its configuration as it fills the active site, displacing solvent molecules in a stepwise fashion. In addition, ZG^{P}LL is able to bind with the Trp 115 water molecule still in place.

The enzyme structure appropriate for binding the transition state of a substrate is not necessarily optimal for binding the ground state conformation. Accordingly, it has been postulated that slow-binding behavior, of the sort exhibited by ZP^{P}LA, might result from a required conformational change on the part of the enzyme or the inhibitor (Wolfenden, 1976; Frieden et al., 1980). Based on tryptophan fluorescence quenching experiments, Kurz et al. (1985) proposed that in the case of adenosine deaminase considerable readjustment of the active site seems required to bind the transition state conformation of the substrate. With respect to the studies presented here, we see no evidence for major rearrangements of the tertiary structure of thermolysin in the active site or elsewhere upon binding ZG^{P}LL or ZP^{P}LA. In the case of thermolysin it appears that the slow-binding phenomenon demonstrated by ZP^{P}LA could be the result of a re-ordering of solvent molecules within the active site rather than any major change in the protein structure.

Steric Hindrance

In considering the different modes of binding of the two inhibitors there is another potential complication that needs to be discussed,

namely steric limitations imposed by the packing of the thermolysin molecules in the crystal. On one hand, the observation that bulky inhibitors such as ZP^{P}LA and phosphoramidon can be bound to crystalline thermolysin shows that the active site is relatively accessible. On the other hand, consideration of the molecular packing within the crystals shows that there is the possibility of steric interference for inhibitors occupying the S_2 and S_3 subsites. These potential steric limitations can be of two types. Either an inhibitor bound to one thermolysin molecule can clash with another thermolysin molecule, or inhibitor molecules bound to two active sites may interfere with each other.

The situation for ZP^{P}LA is shown in Figure 30(b). The closest "intermolecular" approach is 3.5Å and is between the carbobenzoxy group of the inhibitor and Thr 48 of a two-fold-related thermolysin molecule. This distance is just beyond the range of significant van der Waals interactions, although it does not exclude possible indirect effects mediated via bound solvent molecules. However, since there is no direct steric clash between ZP^{P}LA and any other protein or inhibitor molecule in its neighborhood, it is reasonable to assume that the mode of binding of ZP^{P}LA seen in the crystals is very similar to that in solution.

For ZG^{P}LL the potential intermolecular contacts are somewhat different (Fig. 30(a)). Here the carbobenzoxy group of one bound inhibitor comes within 3.5Å of the carbobenzoxy group of another inhibitor molecule bound to the 2-fold-related thermolysin molecule. There are also some intermolecular inhibitor-neighbor protein approaches of 3.8-4.0Å. However, as with ZG^{P}LL, these distances all appear to be

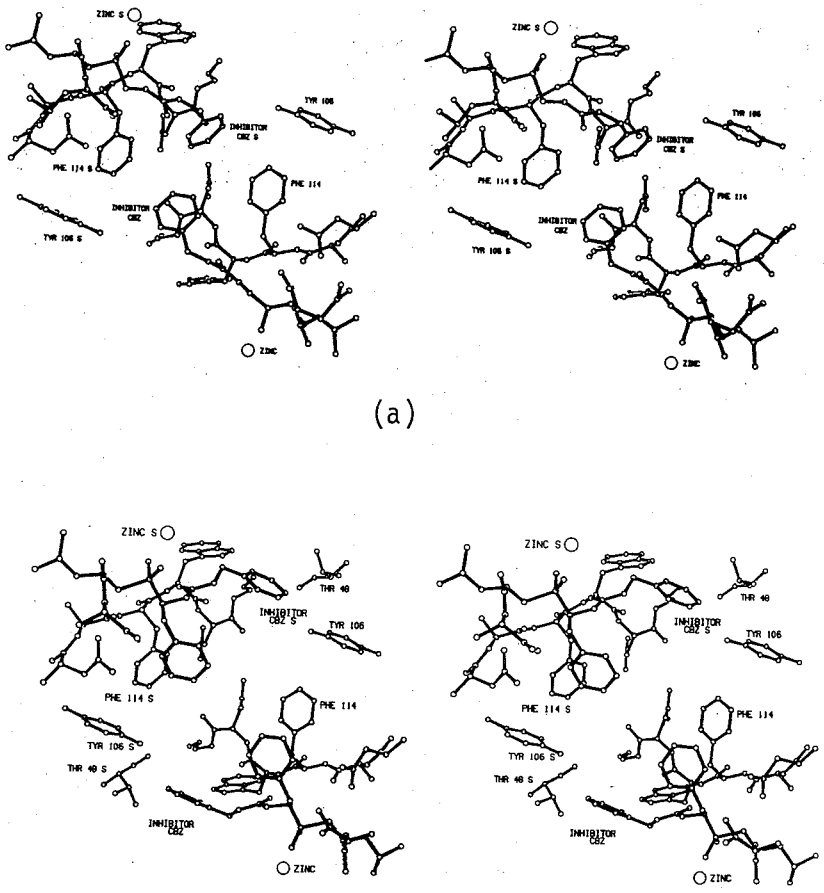


FIGURE 30. Stereo Drawings of Potential Contacts of Inhibitor with Symmetry Related Molecules

Stereo drawings illustrating the potential for steric interference with inhibitors bound in the active site of crystalline thermolysin. Each figure shows one inhibitor molecule (solid bonds) bound in the active site (open bonds) together with a second inhibitor molecule bound to the active site of a neighboring protein molecule. The names of the symmetry-related atoms are followed by an "S". The two active sites and the bound inhibitors are related by a crystallographic two-fold rotational axis that is approximately normal to the plane of the paper. (a) ZG^{P}LL. (b) ZP^{P}LA.

too long to suggest direct intermolecular interactions. Thus, both ZP^{P}LA and ZG^{P}LL come within the vicinity of neighboring molecules in the crystal, but in neither case is there evidence that these neighboring molecules sterically interfere with the preferred mode of binding.

One observation that led us to be concerned about possible steric interference was the difference in density at the phosphorus position in Figures 23(a) and 23(b). As can be seen in the figures, the peak density for ZP^{P}LA is about 60% that for ZG^{P}LL, suggesting that ZP^{P}LA might be bound to the enzyme with less than 100% occupancy. To test this possibility, we refined the occupancy parameters of two models. In the first model the active site contained a molecule of ZP^{P}LA in which the occupancies of all the atoms were constrained to be equal. The second model included, in addition, the solvent molecules which are displaced when the inhibitor is bound. The solvent atoms were constrained to have equal occupancies, but were not tied to the occupancy of the inhibitor. Both calculations lead to the conclusion that ZP^{P}LA occupies the active sites of 80-85% of the thermolysin molecules in the crystal phase. As a control, similar calculations were carried out for ZG^{P}LL and led to an inhibitor occupancy of 100%. The difference in occupancy might explain why the thermal factors of ZP^{P}LA are consistently higher than those of ZG^{P}LL (Table 24). Since the parameters for ZP^{P}LA given in Table 24 were obtained assuming 100% occupancy for the inhibitor, the thermal factor, would tend to increase to compensate for any loss of electron density due to partial occupancy.

APPENDIX A

EFFICIENT CALCULATION OF GRADIENTS BY MEANS OF THE FAST FOURIER TRANSFORM ALGORITHM

Definitions

The following symbols are used in both the chapters and the appendices.

$\underline{p}$	The vector consisting of all refinable parameters in the model.
$\underline{p}_i$	The vector of all parameters of atom i.
$\underline{x}$	The vector of all positional parameters in the model.
$\underline{x}_i$	The vector of the positional parameters of atom i.
x_i, y_i, z_i	The coordinates of atom i.
B_i	The thermal factor for atom i.
O_i	The occupancy of atom i.
f_i	The atomic scattering factor for atom i.
$\underline{s}$	The vector (h,k,l); $s = \sin\theta/\lambda$.
$Q_o(j)$	The observed value for observation j.
$Q_c(j,\underline{p})$	The value calculated for observation j from the parameter p.
$W(j)$	A weighting factor for observation j.
$\lvert F_o(\underline{s}) \rvert$	The observed structure factor for reflection $\underline{s}$.
$\lvert F_c(\underline{s},\underline{p}) \rvert$	The structure factor for reflection $\underline{s}$ calculated from the parameter $\underline{p}$.

T	A Fourier transform.
T^{-1}	An inverse Fourier transform.
$*$	A convolution
$(\underline{x}_i)^*$	The value of the convolution evaluated at $\underline{x}_i$.

Lifchitz Variation of the Algorithm

To refine a structural model against crystallographic data both the function (A.1) and its gradient (A.2) must be evaluated.

$$M(\underline{p}) = \sum_{\underline{s}} W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]^2 \qquad (A.1)$$

$$\frac{dM(\underline{p})}{d\underline{p}} = -2 \sum_{\underline{s}} W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid] \frac{d \mid F_c(\underline{s},\underline{p}) \mid}{d\underline{p}} \qquad (A.2)$$

Because the evaluation of the gradient involves a sum over all reflections for each parameter of the model, the calculation would be very time-consuming if performed as in Equation (A.2).

Agarwal (1978) showed that Equation (A.2) can be expressed as Equations (A.3a-A.3d)

$$\frac{\partial M(\underline{p})}{\partial x_i} = T^{-1}(-2W(\underline{s})(k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid)\exp(i\phi_c(\underline{s},\underline{p}))(2\pi i h)g_i(\underline{s})] \qquad (A.3a)$$

$$\frac{\partial M(\underline{p})}{\partial y_i} = T^{-1}(-2W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))(2\pi i k)g_i(s)] \qquad (A.3b)$$

$$\frac{\partial M(\underline{p})}{\partial z_i} = T^{-1}(-2W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))(2\pi i l)g_i(\underline{s})] \qquad (A.3c)$$

$$\frac{\partial M(\underline{p})}{\partial B_i} = T^{-1}[-2W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))(\frac{-s^2}{4})g_i(\underline{s})] \qquad (A.3d)$$

where $g_i(\underline{s}) = O_i f_i(\underline{s})\exp(B_i s^2/4)$.

This formulation requires a Fourier transform for each parameter and also is impractical. However, Agarwal separated the portions of (A.3) that depend on the atom (subscript i) from the rest of the factors and applied the convolution theorem to arrive at Equations (A.4a-A.4d).

$$\frac{\partial M(\underline{p})}{\partial x_i} = \left(T^{-1}\left[W(\underline{s})\left[k\,|\,F_o(\underline{s})\,|\,-\,|\,F_c(\underline{s},\underline{p})\,|\,\right]\exp(i\phi_c(\underline{s},\underline{p}))(2\pi ih)\right]\right)^*_{(\underline{x}_i)} \quad T^{-1}[-2g(\underline{s})]) \tag{A.4a}$$

$$\frac{\partial M(\underline{p})}{\partial y_i} = \left(T^{-1}\left[W(\underline{s})\left[k\,|\,F_o(\underline{s})\,|\,-\,|\,F_c(\underline{s},\underline{p})\,|\,\right]\exp(i\phi_c(\underline{s},\underline{p}))(2\pi ik)\right]\right)^*_{(\underline{x}_i)} \quad T^{-1}[-2g_i(\underline{s})]) \tag{A.4b}$$

$$\frac{\partial q(\underline{p})}{\partial z_i} = \left(T^{-1}\left[W(\underline{s})\left[k\,|\,F_o(\underline{s})\,|\,-\,|\,F_c(\underline{s},\underline{p})\,|\,\right]\exp(i\phi_c(\underline{s},\underline{p}))(2\pi il)\right]\right)^*_{(\underline{x}_i)} \quad T^{-1}[-2g_i(\underline{s})]) \tag{A.4c}$$

$$\frac{\partial q(\underline{p})}{\partial B_i} = \left(T^{-1}\left[W(\underline{s})\left[k\,|\,F_o(\underline{s})\,|\,-\,|\,F_c(\underline{s},\underline{p})\,|\,\right]\exp(i\phi_c(\underline{s},\underline{p}))\left(\frac{-s^2}{4}\right)\right]\right)^*_{(\underline{x}_i)} \quad T^{-1}[-2g_i(\underline{s})]) \tag{A.4d}$$

If the atomic scattering factor is modeled as a sum of Gaussians the second Fourier transform can be calculated analytically. For a given model, the first transform in each of Equations (A.4a-A.4b) can be calculated with the FFT algorithm. In the form given in Equations (A.4), it is necessary to perform 3 Fourier transforms to determine the positional derivatives.

Lifchitz (see Isaacs, 1982) has pointed out that the gradient can be calculated more efficiently by factoring Equations (A.3) so that the

first Fourier transform is the same in all the equations. The refactoring results in Equations (A.5a-A.5d).

$$\frac{\partial M(\underline{p})}{\partial x_i} = (T^{-1}[W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))]^*_{(\underline{x}_i)} \; T^{-1}[-2(2\pi ih)g_i(\underline{s})]) \qquad (A.5a)$$

$$\frac{\partial M(\underline{p})}{\partial y_i} = (T^{-1}[W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))]^*_{(\underline{x}_i)} \; T^{-1}[-2(2\pi ik)g_i(\underline{s})]) \qquad (A.5b)$$

$$\frac{\partial M(\underline{p})}{\partial z_i} = (T^{-1}[W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))]^*_{(\underline{x}_i)} \; T^{-1}[-2(2\pi il)g_i(\underline{s})]) \qquad (A.5c)$$

$$\frac{\partial M(\underline{p})}{\partial B_i} = (T^{-1}[W(\underline{s})[k \mid F_o(\underline{s}) \mid - \mid F_c(\underline{s},\underline{p}) \mid]\exp(i\phi_c(\underline{s},\underline{p}))]^*_{(\underline{x}_i)} \; T^{-1}[-2(\frac{-s^2}{4})g_i(\underline{s})]) \qquad (A.5d)$$

When using these equations to calculate the gradient of P, only a single FFT is required. The Fourier transforms on the right can still be determined analytically by using the following rules.

$$T^{-1}[e^{-Bs^2}] = (\frac{\pi}{B})^{3/2} \exp(-\pi^2/Br^2) \qquad (A.6)$$

$$T^{-1}[2\pi ihF(\underline{s})] = -\frac{\partial}{\partial x} T^{-1}[F(\underline{s})] \qquad (A.7)$$

$$T^{-1}[s^2 e^{-Bs^2}] = (\frac{\pi}{B})^{3/2} \frac{1}{2B} \exp(-\pi^2 r^2/B)[3 - 2(\frac{\pi^2}{B})r^2] \qquad (A.8)$$

The Reduction of Series Termination Errors

The calculation of the convolutions in Equations (A.5) involves sampling each of the functions at discrete points, multiplying the values point by point, and summing all of the products. The sampling interval required to represent a function with a given accuracy depends on the magnitude of the high resolution components of that function. There is no problem determining the sampling interval for the function on the left of the convolution because it contains no components of resolution higher than the measured data. However, the functions on the extreme right hand side of Equations (A.5) are not of limited resolution and therefore, at least in principle, must be sampled on a very fine grid. This problem existed in the original Equations (A.4) but is much more serious with new form of the Equations (A.5) because the high resolution components are enhanced by the inclusion of the crystallographic indices (h,k,l). This problem is most severe in the calculation of the temperature factor derivatives (Equation A.5d) because of the s^2 factor. A mechanism to allow these functions to be calculated using a somewhat courser grid has been devised by recognizing that the errors introduced by a course grid are fundamentally the same as those encountered in the calculation of structure factors using the FFT method (Ten Eyck, 1977). The solution involves "smearing" or "blurring" the function of interest so that it is sampled by a larger number of grid points. The "smearing" must be compensated elsewhere in the calculation. In this case the compensation is achieved by "sharpening" the difference map. This can be done without introducing

additional errors because no new high resolution terms are introduced into the difference map. The equations used in the program ADERIV,

$$\frac{\partial M(\underline{p})}{\partial x_j} = \left(T^{-1}\left[W(s)\left[k\,|F_o(\underline{s})| - |F_c(\underline{s},\underline{p})|\right]\exp(i\phi_c(\underline{s},\underline{p}))\exp(B^o s^2/4)\right]^{*}_{(\underline{x}_j)}\right.$$
$$\left.T^{-1}\left[-2(2\pi i h)g_j(\underline{s})\exp(-B^o s^2/4)\right]\right) \qquad \text{(A.9a)}$$

$$\frac{\partial M(\underline{p})}{\partial y_j} = \left(T^{-1}\left[W(s)\left[k\,|F_o(\underline{s})| - |F_c(\underline{s},\underline{p})|\right]\exp(i\phi_c(\underline{s},\underline{p}))\exp(B^o s^2/4)\right]^{*}_{(\underline{x}_j)}\right.$$
$$\left.T^{-1}\left[-2(2\pi i k)g_j(\underline{s})\exp(-B^o s^2/4)\right]\right) \qquad \text{(A.9b)}$$

$$\frac{\partial M(\underline{p})}{\partial z_j} = \left(T^{-1}\left[W(s)\left[k\,|F_o(\underline{s})| - |F_c(\underline{s},\underline{p})|\right]\exp(i\phi_c(\underline{s},\underline{p}))\exp(B^o s^2/4)\right]^{*}_{(\underline{x}_j)}\right.$$
$$\left.T^{-1}\left[-2(2\pi i l)g_j(\underline{s})\exp(-B^o s^2/4)\right]\right) \qquad \text{(A.9c)}$$

$$\frac{\partial M(\underline{p})}{\partial B_j} = \left(T^{-1}\left[W(s)\left[k\,|F_o(\underline{s})| - |F_c(\underline{s},\underline{p})|\right]\exp(i\phi_c(\underline{s},\underline{p}))\exp(B^o s^2/4)\right]^{*}_{(\underline{x}_j)}\right.$$
$$\left.T^{-1}\left[-2\left(\frac{-s^2}{4}\right)g_j(\underline{s})\exp(-B^o s^2/4)\right]\right) \qquad \text{(A.9d)}$$

Generalized Uses of the Algorithm

What was not clear in Agarwal's original paper was that this computational short cut can be used in many cases other than the evaluation of Equation (A.2). If the derivation is carried out for the general case we discover that the identity (A.10) holds whenever $E(\underline{s})$ is a symmetric function.

$$\sum_{\underline{s}} E(\underline{s})\frac{\partial |F_c(\underline{s},\underline{p})|}{\partial x_j} = \frac{-1}{2}\left(T^{-1}\left[E(s)\exp(i\phi_c(\underline{s},\underline{p}))\right]^{*}_{(\underline{x}_j)}\right.$$
$$\left.T^{-1}\left[-2(2\pi i h)g_j(\underline{s})\right]\right) \qquad \text{(A.10a)}$$

$$\sum_{\underline{s}} E(\underline{s}) \frac{\partial \left| F_c(\underline{s},\underline{p}) \right|}{\partial y_i} = \frac{-1}{2}\left(T^{-1}[E(s)\exp(i\phi_c(\underline{s},\underline{p}))]^{*}_{(\underline{x}_i)} \; T^{-1}[-2(2\pi ik)g_i(\underline{s})]\right) \qquad \text{(A.10b)}$$

$$\sum_{\underline{s}} E(\underline{s}) \frac{\partial \left| F_c(\underline{s},\underline{p}) \right|}{\partial z_i} = \frac{-1}{2}\left(T^{-1}[E(s)\exp(i\phi_c(\underline{s},\underline{p}))]^{*}_{(\underline{x}_i)} \; T^{-1}[-2(2\pi il)g_i(\underline{s})]\right) \qquad \text{(A.10c)}$$

$$\sum_{\underline{s}} E(\underline{s}) \frac{\partial \left| F_c(\underline{s},\underline{p}) \right|}{\partial B_i} = \frac{-1}{2}\left(T^{-1}[E(s)\exp(i\phi_c(\underline{s},\underline{p}))]^{*}_{(\underline{x}_i)} \; T^{-1}[-2(\tfrac{-s^2}{4})g_i(\underline{s})]\right) \qquad \text{(A.10d)}$$

We can use (A.10) to speed up the calculation of the gradient of almost any function involving structure factors.

Let us develop an example. Suppose that we wish to minimize, not the usual function of the X-ray data (A.1), but the negative of the correlation coefficient $r(\underline{p})$, which can be cast as in Equation (A.11).

$$r'(\underline{p}) = -r(\underline{p}) = \frac{n[\overline{\left| F_o(\underline{s}) \right|}\,\overline{\left| F_c(\underline{s},\underline{p}) \right|} - \overline{\left| F_o(\underline{s}) \right| \left| F_c(\underline{s},\underline{p}) \right|}\,]}{\{[\overline{\left| F_o(\underline{s}) \right|^2} - \overline{\left| F_o(\underline{s}) \right|}^2][\overline{\left| F_c(\underline{s},\underline{p}) \right|^2} - \overline{\left| F_c(\underline{s},\underline{p}) \right|}^2]\}^{1/2}} \qquad \text{(A.11)}$$

The gradient is given by Equation (A.12).

$$\frac{dr'(\underline{p})}{dp} = \frac{r'(\underline{p})}{n}\left[\frac{\left(\overline{\left| F_o(\underline{s}) \right|} \sum_{\underline{s}} \frac{d\left| F_c(\underline{s},\underline{p}) \right|}{d\underline{p}} - \sum_{\underline{s}} \left| F_o(\underline{s}) \right| \frac{d\left| F_c(\underline{s},\underline{p}) \right|}{d\underline{p}}\right)}{\left(\overline{\left| F_o(\underline{s}) \right|}\,\overline{\left| F_c(\underline{s},\underline{p}) \right|} - \overline{\left| F_o(\underline{s}) \right| \left| F_c(\underline{s},\underline{p}) \right|}\right)}\right.$$

$$\left. - \frac{\left(\sum_{\underline{s}} \left| F_c(\underline{s},\underline{p}) \right| \frac{d\left| F_c(\underline{s},\underline{p}) \right|}{d\underline{p}} - \overline{\left| F_c(\underline{s},\underline{p}) \right|} \sum_{\underline{s}} \frac{d\left| F_c(\underline{s},\underline{p}) \right|}{d\underline{p}}\right)}{\left(\overline{\left| F_c(\underline{s},\underline{p}) \right|^2} - \overline{\left| F_c(\underline{s},\underline{p}) \right|}^2\right)}\right] \qquad \text{(A.12)}$$

To calculate this gradient we need a number of means and three complicated summations, (A.13), (A.14), and (A.15).

$$\sum_{\underline{s}} \frac{d\,|F_c(\underline{s},\underline{p})|}{d\underline{p}} \tag{A.13}$$

$$\sum_{\underline{s}} |F_o(\underline{s})| \frac{d\,|F_c(\underline{s},\underline{p})|}{d\underline{p}} \tag{A.14}$$

$$\sum_{\underline{s}} |F_c(\underline{s},\underline{p})| \frac{d\,|F_c(\underline{s},\underline{p})|}{d\underline{p}} \tag{A.15}$$

From the generalized derivation we can see that these three quantities can be calculated from the same convolution, and in fact using the same program as the original calculation but using the three transformations produced by Equations (A.16), (A.17), and (A.18).

$$T^{-1}[\exp(i\phi_c(\underline{s},\underline{p}))] \tag{A.16}$$

$$T^{-1}[|F_o(\underline{s})| \exp(i\phi_c(\underline{s},\underline{p}))] \tag{A.17}$$

$$T^{-1}[|F_c(\underline{s},\underline{p})| \exp(i\phi_c(\underline{s},\underline{p}))] \tag{A.18}$$

Therefore with 3 FFT's we can calculate the required gradient of r'.

This particular function has not been implemented in TNT. To do so would only require the creation of the code to calculate the means, the coefficients for the transformations, and a program which would combine the means with the results of the convolutions to produce the final gradient. To perform refinement a program would have to be written which would calculate r' for any given model. None of these programming tasks are difficult.

APPENDIX B

EVALUATION OF THE GRADIENTS OF THE TERMS NECESSARY FOR THE IMPLEMENTATION OF STEREOCHEMICAL RESTRAINTS

Introduction

In this Appendix we present the equations for the gradients of the stereochemical terms. Most of them are simply derived algebraicly from the quantities definition and no details will be given. Because of the assumptions used in the case of the planarity restraints these equations are derived explicitly.

The stereochemistry terms are of the following form (cf. Equation (2) in the main text).

$$M(\underline{p}) = \sum_b W(b)[Q_o(b) - Q_c(b,\underline{p})]^2 \qquad (B.1)$$

The gradients are of the form

$$\frac{\partial M(\underline{p})}{\partial \underline{x}_i} = -2 \sum_b W(b)[Q_o(b) - Q_c(b,\underline{p})] \frac{\partial Q_c(b,\underline{p})}{\partial \underline{x}_i} \qquad (B.2)$$

Only the portion of (B.2) unique to each type of restraint (i.e. the derivative of the calculated quantity with respect to the parameters of the model) will be listed in each of the following sections. In each section the coordinates are assumed to be expressed in an orthogonal coordinate system.

Bond Lengths

For bond lengths, $Q_c(\underline{x})$ is the distance between two atoms. Call these two atoms i and j.

$$\frac{\partial Q_c(\underline{x})}{\partial x_i} = \frac{1}{Q_c(\underline{x})}(x_i - x_j) \qquad \frac{\partial Q_c(\underline{x})}{\partial x_j} = \frac{-1}{Q_c(\underline{x})}(x_i - x_j)$$

$$\frac{\partial Q_c(\underline{x})}{\partial y_i} = \frac{1}{Q_c(\underline{x})}(y_i - y_j) \qquad \frac{\partial Q_c(\underline{x})}{\partial y_j} = \frac{-1}{Q_c(\underline{x})}(y_i - y_j)$$

$$\frac{\partial Q_c(\underline{x})}{\partial z_i} = \frac{1}{Q_c(\underline{x})}(z_i - z_j) \qquad \frac{\partial Q_c(\underline{x})}{\partial z_j} = \frac{-1}{Q_c(\underline{x})}(z_i - z_j)$$

Bond Angles

For bond angles, $Q_c(\underline{x})$ is the angle defined by three atoms, labeled i, j, k, with j being the central atom.

Let $\underline{a}$ be the vector from atom j to atom i, $\underline{b}$ the vector from atom k to atom j and $\underline{c}$ the vector from atom i to atom k. Then

$$Q_c(\underline{x}) = \cos^{-1}\left[\frac{a^2 + b^2 - c^2}{2ab}\right]$$

[Note: In FORTRAN the argument of the function ACOS is in radians. If $Q_c(\underline{x})$ is to be in degrees a conversion factor must be included.]

$$\frac{\partial Q_c(\underline{x})}{\partial \underline{x}_i} = \frac{-1}{2ab\,|\sin Q_c(\underline{x})|}\left\{\left(1 - \frac{b}{a}\cos Q_c(\underline{x})\right)\frac{da^2}{d\underline{x}_i} + \left(1 - \frac{a}{b}\cos Q_c(\underline{x})\right)\frac{db^2}{d\underline{x}_i} - \frac{dc^2}{d\underline{x}_i}\right\}$$

$$\frac{\partial a^2}{\partial i_w} = 2(i_w - j_w) \qquad \frac{\partial b^2}{\partial i_w} = 0 \qquad \frac{\partial c^2}{\partial i_w} = -2(k_w - i_w)$$

$$\frac{\partial a^2}{\partial j_w} = -2(i_w - j_w) \qquad \frac{\partial b^2}{\partial j_w} = 2(j_w - k_w) \qquad \frac{\partial c^2}{\partial j_w} = 0$$

$$\frac{\partial a^2}{\partial k_w} = 0 \qquad \frac{\partial b^2}{\partial k_w} = -2(j_w - k_w) \qquad \frac{\partial c^2}{\partial k_w} = 2(k_w - i_w)$$

where: $i_1 = X_i \quad j_1 = X_j \quad k_1 = X_k$

$i_2 = Y_i \quad j_2 = Y_j \quad k_2 = Y_k$

$i_3 = Z_i \quad j_3 = Z_j \quad k_3 = Z_k$

Torsion Angles

For torsion angles, $Q_c(\underline{x})$ is the angle defined by the four atoms, labeled i, j, k, l. The quantity is the angle between the normal to the plane defined by atoms i, j, k and the normal to the plane defined by atoms j, k, l.

Let $\underline{a}$ be the vector from atom j to atom i, $\underline{s}$ the vector from atom j to atom k and $\underline{b}$ the vector from atom k to atom l. Also let $\underline{r} = \underline{s} \times \underline{a}$ and $\underline{q} = \underline{s} \times \underline{b}$ where $\underline{r}$ is the normal to the plane containing atoms i, j, k and $\underline{q}$ is the normal to the plane containing atoms j, k, l.

$$Q_c(\underline{x}) = \cos^{-1} \frac{\underline{r} \cdot \underline{q}}{rq}$$

[Note: In FORTRAN the argument of the function ACOS is in radians. If $Q_c(\underline{x})$ is to be in degrees a conversion factor must be included.]

$$\text{Let } \underline{e} = \frac{-1}{rq\,|\sin Q_c(\underline{x})|}\ \{\underline{q} - \cos Q_c(\underline{x})\ \frac{q}{r}\ \underline{r}\}$$

$$\frac{\partial Q_c(\underline{x})}{\partial \underline{x}_i} = \underline{e} \times \underline{s} \qquad \frac{\partial Q_c(\underline{x})}{\partial \underline{x}_l} = \underline{f} \times \underline{s}$$

$$\frac{\partial Q_c(\underline{x})}{\partial \underline{x}_j} = -\frac{\partial Q_c(\underline{x})}{\partial \underline{x}_i} + \underline{e} \times \underline{a} + \underline{f} \times \underline{b} \qquad \frac{\partial Q_c(\underline{x})}{\partial \underline{x}_k} = -\frac{\partial Q_c(\underline{x})}{\partial \underline{x}_l} - \underline{e} \times \underline{a} - \underline{f} \times \underline{b}$$

Planarity

For planarity, $Q_c(\underline{x})$ is defined as the RMS deviation of the atoms from the best fit plane.

Let $\underline{N}$ be the number of atoms in the plane and $\overline{x}$ the center of mass of the atoms. Consider

$$\underline{Q} = \begin{bmatrix} \sum_i (x_i - \overline{x})(x_i - \overline{x}) & \sum (x_i - \overline{x})(y_i - \overline{y}) & \sum (x_i - \overline{x})(z_i - \overline{z}) \\ \sum_i (y_i - \overline{y})(x_i - \overline{x}) & \sum (y_i - \overline{y})(y_i - \overline{y}) & \sum (y_i - \overline{y})(z_i - \overline{z}) \\ \sum_i (z_i - \overline{z})(x_i - \overline{x}) & \sum (z_i - \overline{z})(y_i - \overline{y}) & \sum (z_i - \overline{z})(z_i - \overline{z}) \end{bmatrix}$$

$\underline{Q}$ is the moments matrix for the atoms of the plane. The eigenvectors of $\underline{Q}$ point along the directions of the principle axes of rotation of this group of atoms. The eigenvalues of $\underline{Q}$ are inversely related to the moments of inertia of rotation about the axis defined by the corresponding eigenvector. The axis of rotation with the largest moment (smallest eigenvector) is defined as the normal to the best plane for these atoms.

Let u be the smallest eigenvalue of $\underline{Q}$, $\underline{n}$ the eigenvector of $\underline{Q}$ corresponding to u and $\underline{m}$ ($=\underline{n}/n$) be the normalized eigenvector. Then the

rms deviation of the atoms from planarity is

$$Q_c(\underline{x}) = \left[\sum_i \frac{(\underline{m}\cdot(\underline{x}_i - \overline{\underline{x}}))^2}{N}\right]^{1/2}$$

$$\frac{dQ_c(\underline{x})}{d\underline{x}_i} = \frac{(\underline{m}\cdot(\underline{x}_i - \overline{\underline{x}}))}{NQ_c(\underline{x})}\left[(1 - \frac{1}{N})\underline{m} + (\underline{x}_i - \overline{\underline{x}})\cdot\frac{d\underline{m}}{d\underline{x}_i}\right]$$

$$\frac{d\underline{m}}{d\underline{x}_i} = \frac{1}{n}\frac{d\underline{n}}{d\underline{x}_i} - \frac{\underline{n}}{n^3}\Lambda\left(\underline{n}\cdot\frac{d\underline{n}}{d\underline{x}_i}\right)$$

where Λ is the outer product. It is defined, when $\underline{a}$ and $\underline{b}$ are column vectors, as $\underline{a}\ \Lambda\ \underline{b} = \underline{a} \cdot \underline{b}^T$.

The calculation of the derivative of the eigenvector with respect to the position of an atom is difficult because eigenvectors are usually determined algorithmically. There is no equation which expresses the components of the eigenvector of a matrix as a function of the components of that matrix. However, if one assumes that the off-diagonal elements of $\underline{Q}$ are non-zero one can derive an equation for the eigenvector.

$$n_1 = Q_{31}\,Q_{32} + Q_{21}\,(u - Q_{33})$$
$$n_2 = Q_{31}\,Q_{31} + (u - Q_{33})(u - Q_{11})$$
$$n_3 = Q_{31}\,Q_{21} + Q_{32}\,(u - Q_{11})$$

The derivatives of n_1, n_2 and n_3 are simple to derive in terms of the derivatives of the elements of $\underline{Q}$. Because of the complexity of its derivative we have made the assumption that the eigenvalue remains constant during refinement. The assumption that the off-diagonal elements of $\underline{Q}$ are non-zero make the gradient calculation sensitive to

the orientation of the plane. In the program which performs these calculations the problems which might arise are ignored. It is presumed that, if, by chance, the plane lies in a special orientation the movement resulting from the first cycle of refinement will cause it to be displaced and subsequent refinement will function normally.

$$\frac{dQ}{dx_i} = \left(1 - \frac{1}{N}\right) \begin{bmatrix} 2(x_i - \bar{x}) & (y_i - \bar{y}) & (z_i - \bar{z}) \\ (y_i - \bar{y}) & 0 & 0 \\ (z_i - \bar{z}) & 0 & 0 \end{bmatrix}$$

$$\frac{dQ}{dy_i} = \left(1 - \frac{1}{N}\right) \begin{bmatrix} 0 & (x_i - \bar{x}) & 0 \\ (x_i - \bar{x}) & 2(y_i - \bar{y}) & (z_i - \bar{z}) \\ 0 & (z_i - z) & 0 \end{bmatrix}$$

$$\frac{dQ}{dz_i} = \left(1 - \frac{1}{N}\right) \begin{bmatrix} 0 & 0 & (x_i - \bar{x}) \\ 0 & 0 & (x_i - \bar{y}) \\ (x_i - \bar{x}) & (y_i - \bar{y}) & 2(z_i - \bar{z}) \end{bmatrix}$$

APPENDIX C

SECONDARY STRUCTURE OF THE BACTERIOCHLOROPHYLL PROTEIN

The high resolution model of the bacteriochlorophyll containing protein can be used to identify the residues that occur within regular secondary structure and the interactions that occur within and at the ends of each element of structure. This Appendix itemizes all of the secondary structure elements of the protein. Each element of secondary structure (β, α, or turn) is given a number according to whether it is the first, second, third, etc., element of that type, counting from the N-terminus. A list of all the secondary structure elements is given in Table 28.

It is clear from the Ramachandran plot (Fig. 6) that the Bchl protein is composed primarily of β-sheet. However there are, in fact, only two β-sheets in the structure, with one containing only four strands. The other β-sheet contains 16 strands, many of which are 15 or so amino acids in length. Also contained within the structure of the Bchl protein monomer are 8 α-helices ranging in size from 5 to 15 residues and 11 β bends of various types.

The Greater Sheet

The greater sheet is the large β-sheet that covers the outside of the trimer. It has two parts: the "top half", which is almost flat, and the "bottom half", where each strand is bent by about 90°. The

TABLE 28. Secondary Structure Elements of the Bacteriochlorophyll Protein

Structure element	Residues	Structure element	Residues
Beta-1	3- 14	Alpha-5	225-230
Beta-2	20- 29	Alpha-6	230-234
Beta-3	39- 49	Beta-11	244-253
Beta-4	57- 68	Beta-12	256-265
Beta-5	71- 85	Beta-13	268-272
Beta-6	88-101	Beta-14	276-280
Beta-7	103-118	Col-1	278-283
Alpha-1	121-128	Alpha-7	284-294
Beta-8	135-146	Beta-15	299-309
Alpha-2	150-164	Beta-16	314-319
Alpha-3	169-179	Beta-17	324-326
Alpha-4	183-190	Beta-18	328-333
Beta-9	191-205	Alpha-8	335-346
Beta-10	109-222	Beta-19	352-356

greater sheet is composed of the β strands 1 through 8, 8b, 9 through 12, 15, 17, and 18.

The first amino acid of the present model (Residue #3) is the first amino acid of β strand Beta-1. Any amino acids that might come before this residue could not be an extension of this strand because the amino acid is located at the edge of the major sheet and at a monomer-monomer interface. The only direction the chain could go is up, away from the surface of the trimer. Beta-1 contains 12 amino acids and ends with a connection to Beta-2 via a loop of 7 amino acids, with Turn-1 nested in the middle of the loop. The residues of the loop follow roughly the pattern of antiparallel strands of β-sheet but form none of requisite H-bonds.

Beta-2 is terminated on its N-terminal side with an H-bond between 19:O and 268:N (19:O designates the carbonyl oxygen of residue 19; 269:N is the peptide nitrogen of residue 269, etc.). The strand is antiparallel to Beta-1 and continues until Beta-1 stops. This strand terminates with a possible γ turn. A H-bond connects 29:O and 31:N, however there are peaks on the difference map in this region and the side chain of residue 30 is clearly incorrect. Beta-2 is connected to Beta-3 by a 10 residue stretch with a +3 topological connection.

Beta-3 is bounded on both ends by proline residues. The strand ends in some form of a tight turn but one residue past the terminating proline the interpretation of main chain density is uncertain and the chain is broken in the model.

Beta-4 begins at the break mentioned above, continues for 12 residues, and is terminated at the edge of the sheet with Turn-2.

Beta-5 is a 14 residue strand that ends where the great fold of the major β-sheet begins. The last three residues of the strand (82, 84, 85) extend beyond Beta-4. Residue 86 is in a β-bulge conformation with a H-bond between 86:N and 89:O (3.11Å) (89:O is the β mate of 85:N). Residues 86, 87, 88, and 89 form a β-turn (Turn 2a) with no H-bond between 86:O and 89:N. 88 is the first residue of Beta-6.

Beta-6 crosses the sheet with 14 amino acids and ends with Turn-3 (Type II'). The nitrogen atom of the central peptide bond of the peptide (102:N) is H-bonded (2.73Å) to 127:ND2 of the neighboring monomer.

Beta-7 recrosses the sheet in 16 residues and is the longest strand in the structure. It is also the most curved strand, with a bend of approximately 90° spread over its length. The strand three strands back (Beta-4) was almost straight. The curvature is due to distortions of Beta-5, Beta-6, and Beta-7. Each successive pair of strands (Beta-3,4, Beta-5,6, and Beta-7) begins in register with its predecessor but is 2 residues longer.

Beta-7 is connected to Beta-8 by a long loop of 16 residues that contains helix Alpha-1. This loop is the region of one monomer that interacts most extensively with the next monomer of the trimer. The loop extends well away from the rest of the monomer and at its most distal portion becomes Alpha-1. The loop then returns to the monomer and begins Beta-8.

Beta-8 starts in the same location as Beta-9 and Beta-10 but there are no clear indications as to what interaction might have caused this. The strand crosses the sheet with 12 residues and ends at Proline 147

which is followed by Turn-3a, a γ turn involving residues 147, 148, and 149.

Connecting Beta-8 and Beta-8b is a long loop of 36 residues, which contains three α-helices (Alpha-2, Alpha-3 and Alpha-4), Turn-4 (Type II'), and a break in the model between Alpha-2 and Alpha-3. The connection has a topological number +3.

Beta-8b is a very short strand composed of two residues which are H-bonded to Beta-9. Beta-8b is unusual because it is entirely contained within Alpha-4. Beta-8b terminates with Turn-4b (which is equivalent to Alpha-4 terminating with a 3_{10} H-bond). Beta-4 is the last strand on this end of the Greater Sheet.

Beta-9 begins with Turn-4b. The H-bond of Turn-4b (between atom 191:N and 188:O) is the first H-bond of Beta-9. This is also the 3_{10} H-bond that terminates Alpha-4 (or one could say that this residue terminates Beta-9 and starts Alpha-4).

In two places residues are bulged out of the β sheet H-bonding pattern. The first instance is at residue 196 (PRO). This structure begins with the peptide plane between 195 (SER) and 196 being rotated approximately 90° from the normal H-bonding position. The β-sheet H-bond to the other strand (Beta-10) that is lost because of this distortion is replaced by an H-bond between 195:OG and the nitrogen on the other strand (2.85Å). The proline side chain of 196 is on the same side of the sheet as the serine side chain of 195. The peptide bond between 196 and 197 is also rotated well away from the standard β sheet position so that 196:O also H-bonds to 195:OG (2.90Å) and 197:N makes a

H-bond to 171:O (2.88Å). Residue 171 is in the middle of Alpha-3 and the helix compensates for this peptide bond being turned out by forming a single 3_{10} H-bond near this location. The next H-bond of the β sheet, between 198:N and 216:O, is unusually long (3.10Å) reflecting the strain imposed by this conformation.

The second bulge is near residue 200 (ASX). The main chain nitrogens of both residues 200 and 201 form H-bonds to 214:O (2.56Å and 3.12Å). This is a classic β bulge with no protein on the far side of the strand with which it may interact.

The connection between Beta-9 and Beta-10 is short and has a topology of -1 but there is a break here and little else can be said with certainty.

Beta-10 begins with a possible proline (209), continues between Beta-9 and Beta-8 for a length of 14 residues, and ends for no clear reason.

The connection between Beta-10 and Beta-11 is 21 residues long and follows a reasonably straight but contorted path. The topology of the connection is -7. This loop contains Alpha-5, Alpha-6, Turn-5, and Turn-6.

The first residue of Beta-11 is residue 244 and is directly preceded by a proline. This strand is in the flat portion of the sheet and is only 10 residues long. Beta-11 is connected to Beta-12 by a topologically -1 short turn whose cross H-bond is formed by 253:OD2 and 256:N (3.00Å). 253:OD2 also forms a bond with 255:N (2.84Å).

Beta-12 continues back for 10 residues and is terminated by a proline at residue 266. Beta-12 connects to Beta-15 by a loop of 33 residues. The loop contains two strands of the lesser sheet: Beta-13 and Beta-14, Alpha-7, and a stretch of collagen-like helix named Col-1. The topology of this connection is -3x.

Beta-15 begins next to the N-terminus, and in this vicinity it forms one edge of the greater β sheet. Strand Beta-1 and Beta-15 are parallel strands and are the only two parallel strands in the structure. Beta-15 is terminated by Proline 310 and is connected to Beta-17 by a 14-residue loop of topology -2x containing Turn-9, and Beta-16 of the Lesser Sheet.

Beta-17 is a short strand of 3 residues located at the very "top" end of the Greater Sheet. The first H-bond of this strand (between 324:N and 331:O (of Beta-18)) is weak, being only 3.13Å in length. This is because 331:O also forms an H-bond with 323:N which is very strong (2.36Å).

Turn-10 connects Beta-17 to Beta-18. Beta-18 is 6-residues long and has no clear reason for stopping.

The Lesser Sheet

The Lesser β Sheet of the Bchl protein is a small antiparallel β sheet composed of four strands. The four strands are named Beta-13, Beta-14, Beta-16, and Beta-19. These strands, together with Alpha-8, form the back side of the sandwich with the "upper" part of the Greater Sheet forming the front part and Ring IV wedged between. The Lesser

Sheet has a mild twist, with Beta-16 inclined about 30° from Beta-13 and Beta-14.

Beta-13 begins shortly after Beta-12 ends with a proline. Three residues have their main chain nitrogen atom within H-bonding distance of 19:0 (267-3.24Å, 268-2.90Å, 269-3.20Å) but only the third residue's interaction has reasonable geometry for an H-bond. Residues 268 and 269 are the first residues of this short strand of 7 residues. Between the final residue, 272 and Turn-8 (either Type I' or III') lies residue 273. This residue makes no H-bonds in the present model but it, and the following two residues, are not well determined in the map and could be misplaced. There is no cross H-bond in Turn-8 and that too could be an error. The topology of the turn is -1.

Beta-14 is a 5-residue strand which ends in a distorted β sheet conformation. The last three residues of this strand are also in a region of collagen-like helix called Col-1. This 6-residue section is in a highly hydrated section of the molecule where the main chain hydrogen bonding groups form bonds with the bound solvent. The connection between Beta-14 and Beta-16 has a topological connection of type 2x.

Beta-16 is at the topmost extreme of the sheet and is 6 residues long, but residue 316 is pushed out in a classic β bulge. The strand begins with Turn-9 and ends with Proline 320. Beta-16 connects to its neighbor strand, Beta-19 with a rather involved connection containing Beta-17, Beta-18, Alpha-8, Turn-10, and Turn-11. This connection has a topological number 1.

Beta-19 is a strand that is also 6 residues long. The backbone continues in an extended conformation beyond the hydrogen-bonded sheet region to reach the C-terminus.

α-Helices

There are 8 α-helices in the present model. In the initial unrefined model there were reported to be only seven helices. One of the helices of the early model is now considered to be the two helices Alpha-5 and Alpha-6. Both classifications can be justified in that this could be viewed either as a single helix with a very large kink in its middle or as two helices very close together in the chain.

Alpha-1 is a helix of 8 residues containined within the large loop that extends out to the next monomer of the trimer. The N-terminus is capped by residue 121. The atom 121:O forms an H-bond with the residue one would expect in an α-helix: 125:N (3.05Å) and also forms a weak 3_{10} bond to 124:N (3.13Å). This helix has several interactions at its terminus. First, oxygen 123:O does not bond to 127:N as it would in a normal helix but bonds to atom 127:OD1 (2.99Å). This interaction frees 127:N to make a 3_{10}-type H-bond to 124:O (2.96Å) which also makes a normal α-helix H-bond to 128:N (2.96Å).

Alpha-2 is the long helix (15 residues) that seems to protrude from the "bottom" of the molecule. The N-terminal side of the helix is terminated with an H-bond between 150:OD1 and 153:N (3.12Å). The C-terminal end of this helix is at the site of the second break in the main chain of the model.

The first atom after the break is within Alpha-3. This helix (the visible region at least) begins with residue 169 and extends to 179. There is no density prior to residue 169 and there are very few neighboring residues. In the middle of the helix there is a single 3_{10} helix H-bond where the normal α-helix H-bonding pattern is broken to exclude a single carbonyl oxygen. This oxygen (171:O) binds to the peptide bond between 196 and 197 which is within a distorted β bulge. Oxygen 173:O is also excluded from the helix but forms no H-bonds. Because of the distortion of this helix, the helical axis is bent by about 20°.

The first carbonyl oxygen that does not H-bond within the helix is 176:O and it makes no H-bonds at all to other atoms in the model. 177:O forms a bond with 135:OE1 of the neighboring monomer with a distance of 3.84Å. The helix terminates with a γ H-bond (Turn-3b) between 178:O and 180:N (3.20Å).

Between Alpha-3 and Alpha-4 are three residues including both Turn-3b and Turn-4. This stretch of residues lies near the monomer-monomer interface and interacts both with phytol tails in the interior and residues of the neighboring monomer.

Residue 183 is the first residue of Alpha-4 to have α-helical phi-psi angles but atom 186:O is the first to make an α-helical H-bond (3.16Å). The pitch of the helix before 186 is too steep to allow the oxygen atoms to come close enough to the nitrogen atoms to make the required bonds but other bonds are made. The termination of Alpha-4 is unusual. The C-terminal end of the helix is capped with a 3_{10}-type H-bond that doubles as a β turn (Turn-4b). Two of the residues

preceding this turn (187, 188) form a small β strand (Beta-8b) that is contained entirely within the helix.

Residue 188 (ASX) is one of the two residues whose side chain is very near both of its symmetry related mates in the Bchl-protein trimer. The side chain of 188 is about 4.5Å from both of symmetry mates and does not appear to directly interact with them. There is a cavity that extends along the trimer axis and narrows here, but never closes completely.

Alpha-5 begins with residue 225 and extends until residue 230. This short helix has only two α-helical H-bonds (225:O - 229:N and 226:O - 230:N). The helix ends with two H-bonds outside the helix. First, atom 227:O bonds to 138:OE2 with a distance of 2.95Å and then 228:O bonds to 117:OH (3.04Å).

Alpha-6 is even shorter than Alpha-5, being only 5 residues long. Residue 230 is within both Alpha-5 and Alpha-6 but the axes of the two helices are about 90° apart. Alpha-6 has but one helical H-bond. This bond connects residue 230 with 234. 231:O bonds to 90:NH1 (2.86Å) while 232:O makes a 3_{10} link to 235:N and 233:O bonds to SOL1:4OS (2.86Å). 234:O is the ligand of the magnesium of Ring V and bonds with a length of 2.03Å. The helical axis does not pass through the ring and forms an angle of about 30° with the normal to the ring plane.

Alpha-7 begins with residue 284 and ends with residue 309. The peptide bond just before the helix begins (between residues 283 and 284) forms an H-bond bridge between the side chains of residues 282 (HIS) and 226 (ASX). Residue 282 is the ligand to the magnesium in Ring IV. If

283:O were moved into the proper position to H-bond to 287:N, the two proline residues 283 and 286 would come too close together. The helix terminates when 288:O H-bonds to SOL2:160 (2.42Å) instead of 292:N. 292:N compensates by making a 3_{10} bond to 289:O (2.89Å). Atom 290:O makes two H-bonds, one with 293:N (2.79Å) and the other with 294:N (2.71Å). The first bond is a 3_{10} type H-bond while the second is a normal α-helix H-bond.

Alpha-8 begins with a 3_{10}-type H-bond that is labeled Turn-11. The helix extends from residue 335 for 12 residues and ends with residue 346. Atom 334:O makes the cross H-bond of Turn-11. 335:O makes two H-bonds. The first is with 338:N (3.01Å) while the second is with 339:N (3.06Å). The helix ends where atom 346:N makes a weak bifurcated H-bond to 342:O (3.26Å) and 343:O (3.17Å).

BIBLIOGRAPHY

Agarwal, R.C. (1978) Acta Cryst. A34, 791-809.

Alber, T., Grütter, M.G., Gray, T.M., Wozniak, J.A., Weaver, L.H., Chen, B-L., Baker, E.N., & Matthews, B.W. (1986) UCLA Symposia on Molecular and Cellular Biology, "Protein Structure, Folding and Design", Alan R. Liss, Inc., in press.

Anderson, W.F., Ohlendorf, D.H., Takeda, Y., & Matthews, B.W. (1981) Nature 290, 754-758.

Antonov, V.K., Ginodman, L.M., Rumsh, L.D., Kapitannikov, Y.V., Barshevskaya, T.N., Yavashev, L.P., Gurova, A.G., & Volkova, L.I. (1981) Eur. J. Biochem. 117, 195-200.

Baker, E.N., & Hubbard, R.E. (1984) Prog. Biophys. Molec. Biol. 44, 97-179.

Bartlett, P.A., & Marlowe, C.K. (1983) Biochemistry 22, 4618-4624.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B. Jr., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) J. Mol. Biol. 112, 535-542.

Bolognesi, M.C., & Matthews, B.W. (1979) J. Biol. Chem. 254, 634-639.

Bowen, H.J.M., Donohue, J., Jenkin, D.G., Kennard, O., Wheatley, P.J., & Whiffen, D.H. (1958) in Tables of Interatomic Distances and Configuration in Molecules and Ions (Mitchell, A.D., & Cross, L.C., Eds.) The Chemical Society, London.

Chow, H.-C., Serlin, R. & Strouse, C.E. (1975) J. Amer. Chem. Soc. 97, 7230-7237.

Colman, P.M., Jansonius, J.N. & Matthews, B.W. (1972) J. Mol. Biol. 70, 701-724.

Cushman, D.W., Cheung, H.S., Sabo, E.F., & Ondetti, M.A. (1977) Biochemistry 16, 5484-5491.

Deisenhofer, J., Epp, O., Huber, R., & Michel, H. (1984) J. Mol. Biol. 180, 385-398.

Diamond, R. (1966) Acta Cryst. 21, 253-266.

Diamond, R. (1971) Acta Cryst. A27, 436-452.

Dodson, E.J., Isaacs, N.W., & Rollett, J.S. (1976) Acta Cryst. 32, 311-315.

Fenna, R.E., & Matthews, B.W. (1975) Nature 258, 573-577.

Fenna, R.E., TenEyck, L.F., & Matthews, B.W. (1977) Biochem. Biophys. Res. Commun. 75, 751-756.

Fletcher, R., & Reeves, C.M. (1964) Comput. J. 7. 149-154.

Forsyth, J.B., & Wells, M. (1959) Acta Cryst. 12, 412-415.

Freer, S.T., Alden, R.A., Carter, C.W. Jr., & Kraut, J. (1975) J. Biol. Chem. 250, 46-54.

Frieden, C., Kurz, L.C., & Gilbert, H.R. (1980) Biochemistry 19, 5303-5309.

Galardy, R.E. (1982) Biochemistry 21, 5777-5781.

Galardy, R.E., Kontoyiannidou-Ostrem, Y., & Kortylewicz, Z.P. (1983) Biochemistry 22, 1990-1995.

Gardell, S.J., Craik, C.S., Hilvert, D., Urdea, M.S., & Rutter, W.J. (1985) Nature 317, 551-554.

Grütter, M.G., Weaver, L.H., Gray, T., & Matthews, B.W. (1983) in Bacteriophage T4 (Matthews, C.K., Kutter, E., Mosig, G., & Berget, P., Eds.) pp. 356-360, American Society for Microbiology, Washington, D.C.

Hangauer, D.G., Monzingo, A.F., & Matthews, B.W. (1984) Biochemistry 23, 5730-5741.

Hendrickson, W.A., & Konnert, J.H. (1980) in Computing in Crystallography (Diamond, R., Ramaseshan, S., & Venkatesan, K., Eds.) pp. 13.01-13.25, Indian Academy of Sciences, Bangalore.

Hermans, J. Jr., & McQueen, J.E. Jr. (1974) Acta Cryst. A30, 730-739.

Holmes, M.A., & Matthews, B.W. (1981) Biochemistry 20, 6912-6920.

Holmes, M.A., & Matthews, B.W. (1982) J. Mol. Biol. 160, 623-639.

Holmes, M.A., Tronrud, D.E., & Matthews, B.W. (1983) Biochemistry 22, 236-240.

Holmquist, B. (1977) Biochemistry 16, 4591-4594.

Holmquist, B., & Vallee, B.L. (1979) Proc. Natl. Acad. Sci. USA 76, 6216-6220.

International Tables for Crystallography, Vol. II (1962) The Kynoch Press, Birmingham, England.

Isaacs, N. (1982) in Computational Crystallography (Sayre, D., Ed.) Clarendon Press, Oxford, pp. 398-408.

Jack, A., & Levitt, M. (1978) Acta Crystallogr. A34, 931-935.

Jacobsen, N.E., & Bartlett, P.A. (1981) J. Amer. Chem. Soc. 103, 654-657.

Jeffrey, G.A., & Maluszynska, H. (1982) Int. J. Biol. Macromolecules 4, 173-185.

Jones, T.A. (1982) in Crystallographic Computing (Sayre, D., Ed) pp. 303-317, Oxford University Press.

Jones, T.A., & Liljas, L. (1984) Acta Cryst. A40, 50-57.

Kabsch, W. (1976) Acta Crystallogr. A32, 922-923.

Kam, C-M., Nishino, N., & Powers, J.C. (1979) Biochemistry 18, 3032-3038.

Kester, W.R., & Matthews, B.W. (1977) Biochemistry 16, 2506-2516.

Kester, W.R., & Matthews, B.W. (1977b) J. Biol. Chem. 252, 7704-7710.

Kojima, T., Yamane, T., & Ashida, T. (1978) Acta Crystallogr. B34, 2896-2898.

Komiyama, T., Suda, H., Aoyagi, T., Takeuchi, T., & Umezawa, S. (1975) Arch. Biochem. Biophys. 171, 727-731.

Kratky, C. & Dunitz, J.D. (1975) Acta Crystallogr. B31, 1586-1589.

Kunugi, S., Hirohara, H., & Ise, N. (1982) Eur. J. Biochem. 124, 157-163.

Kuo, L.C., & Makinen, M.W. (1982) J. Biol. Chem. 254, 24-27.

Kurz, L.C., LaZard, D., & Frieden, C. (1985) Biochemistry 24, 1342-1346.

Lutz, M., Hoff, A.J. & Brehamet, L. (1982) Biochim. Biophys. Acta 679, 331-341.

Matthews, B.W. (1977) in The Proteins (Neurath, H. & Hill, R.L., Eds.) Vol. 3, 3rd Ed., pp. 403-590, Academic Press, New York.

Matthews, B.W. (1982) in Lipid-Protein Interactions, Vol. 1 (Jost, P., & Griffith, O.H., Eds.) pp. 1-23, John Wiley.

Matthews, B.W., Fenna, R.E., Bolognesi, M.C., Schmid, M.F. & Olson, J.M. (1979) J. Mol. Biol. 131, 259-285.

Matthews, B.W., Weaver, L.H., & Kester, W.R. (1974) J. Biol. Chem. 249, 8030-8044.

Maycock, A.L., Desousa, D.M., Payne, L.G., ten Broeke, J., Wu, M.T., & Patchett, A.A. (1981) Biochem. Biophys. Res. Commun. 102, 963-969.

Molnar, C.E., Barry, C.D., & Rosenberger, F.U. (1976) Tech. Memo. No. 229, Computer Systems Laboratory, Washington University, St. Louis.

Monzingo, A.F., & Matthews, B.W. (1982) Biochemistry 21, 3390-3394.

Monzingo, A.F., & Matthews, B.W. (1984) Biochemistry 23, 5724-5729.

Morihara, K., & Tsuzuki, H. (1970) Eur. J. Biochem. 15, 374-380.

Nishino, N., & Powers, J.C. (1979) J. Biol. Chem. 255, 3482-3486.

Olson, J.M. (1971) in Photosynthesis Part A (San Pietro, A., Ed.) (Vol. 23 of Methods of Enzymology), pp. 636-644, Academic Press, New York.

Olson, J.M. (1978) in The Photosynthetic Bacteria (Clayton, R.K. & Sistrom, W.R., Eds.) pp. 161-197, Plenum Press, New York.

Olson, J.M., Shaw, E.K., & Engleberger, F.M. (1976) Biochem. J. 159, 769-774.

Ondetti, M.A., Condon, M.W., Reid, J., Sabo, E.F., Cheung, H.S., & Cushman, D.W. (1979) Biochemistry 18, 4340-4347.

Ondetti, M.A., Rubin, B., & Cushman, D.E. (1977) Science (Washington, D.C.) 196, 441-444.

Palmer, A.R., Ellis, P.D., & Wolfenden, R. (1982) Biochemistry 21, 5056-5059.

Pangburn, M.K., & Walsh, K.A. (1975) Biochemistry 14, 4050-4054.

Pascher, I. (1976) Biochim. Biophys. Acta 455, 433-451.

Pascher, I., & Sondell, S. (1977) Chem. Phys. Lipids 20, 175-191.

Pauling, L. (1946) Chem. Eng. News 24, 1375-1377.

Pearson, R.H., & Pascher, I. (1979) Nature (London) 281, 499-501.

Perlstein, R.M., & Hemenger, R.P. (1978) Proc. Natl. Acad. Sci. USA 75, 4920-4924.

Ramachandran, G.N., Ramakrishnan, C., & Sasisekharan, V. (1963) J. Mol. Biol. 7, 95-99.

Ramachandran, G.N., & Sasisekharan, V. (1968) Advan. Prot. Chem. 23, 283-438.

Richards, F.M. (1968) J. Mol. Biol. 37, 225-230.

Robertus, J.D., Kraut, J., Alden, R.A., & Birktoft, J.J. (1972) Biochemistry 11, 4293-4303.

Rose, G.D., Young, W.B., & Gierasch, L.M. (1983) Nature 304, 655-657.

Rossmann, M.G. (1979) J. Appl. Crystallogr. 12, 225-238.

Schechter, I., & Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157-162.

Schmid, M.F., Tronrud, D.E., & Matthews, B.W. (1983) Chemica Scripta 21, 67-69.

Schmid, M.F., Weaver, L.H., Holmes, M.A., Grutter, M.G., Ohlendorf, D.H., Reynolds, R.A., Remington, S.J., & Matthews, B.W. (1981) Acta Crystallogr. A37, 701-710.

Soman, K.V., & Ramakrishnan, C. (1983) J. Mol. Biol. 170, 1045-1048.

Stout, G.H., & Jensen, L.H. (1968) X-ray Structure Determination: A Practical Guide, Macmillan Publishing Co., Inc., London, England.

Strouse, C.E. (1974) Proc. Natl. Acad. Sci. USA 71, 325-328.

Suda, H. Aoyagi, T., Takeuchi, T., & Umezawa, H. (1973) J. Antibiot. (Tokyo) 26, 621-623.

Sussmann, J.L., Holbrook, S.R., Church, J.M., & Kim, S-H. (1977) Acta Cryst. A33, 800-804.

Ten Eyck, L.F. (1977) Acta Cryst. A33, 486-492.

Ten Eyck, L.F., Weaver, L.H., & Matthews, B.W. (1976) Acta Cryst. A32, 349-350.

Thorsett, E.D., Harris, E.E., Peterson, E.R., Greenlee, W.J., Patchett, A.A., Ulm, E.H., & Vassil, T.C. (1982) Proc. Natl. Acad. Sci. USA 79, 2176-2180.

Venkatachalam, C.M. (1968) Biopolymers 6, 1425-1436.

Vijayan, M. (1976) in CRC Handbook of Biochemistry and Molecular Biology (Fasman, G.D., Ed.) 3rd Edition, Proteins, Vol. II, pp. 742-759, CRC Press, Cleveland.

Warme, P.K., Gō, W., & Scheraga, H.A. (1972) J. Comput. Phys. 9, 303-317.

Watenpaugh, K.D., Sieker, L.C., Herriott, J.R., & Jensen, L.H. (1973) Acta Cryst. B29, 943-956.

Weaver, L.H., Kester, W.R., & Matthews, B.W. (1977) J. Mol. Biol. 114, 119-132.

Wlodawer, A. and Hendrickson, W.A. (1982) Acta Cryst. A38, 239-247.

Wlodawer, A., Walter, J., Huber, R., & Sjölin, L. (1984) J. Mol. Biol. 180, 301-329.

Wolfenden, R. (1976) Ann. Rev. Biophys. Bioeng. 5, 271-306.