CRYSTAL STRUCTURE AND HYDROXYAPATITE BINDING OF PORCINE OSTEOCALCIN

By

Quyen Quoc Hoang

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

McMaster University

STRUCTURE AND FUNCTION OF OSTEOCALCIN

DOCTOR OF PHILOSOPHY (2003)   McMaster University

(Biochemistry)   Hamilton, Ontario

TITLE: Crystal structure and hydroxyapatite binding of osteocalcin from porcine.

AUTHOR: Quyen Quoc Hoang, B. Sc. (McMaster University)

SUPERVISOR: Dr. Daniel S. C. Yang

NUMBER OF PAGES: xiii, 138

Abstract

Osteocalcin is a bone specific protein produced by osteoblast. The strong association of osteocalcin with bone rendered it a principal biological marker for bone metabolism and the most characterized among non-collagenous bone matrix proteins. To gain insight into the function of osteocalcin, the crystal structure of osteocalcin has been determined to 2.0 Å resolution. Contrary to a previous hypothesis that osteocalcin is largely disordered; the crystal structure of osteocalcin is tightly folded. One surface of the structure is relative flat and consists of only negatively charged residues. The arrangement of carboxyl groups on the γ-carboxylated glutamic acid residues closely mimic the arrangement of phosphate groups in hydroxyapatite, and they chelate five calcium atoms in the dimer interface with near perfect coordination geometry. When docked to the surface of hydroxyapatite, this surface complimented near perfectly the prism face of hydroxyapatite. The excellent complimentarity clearly indicates that osteocalcin can bind to hydroxyapatite with high specificity, it seems reasonable that the presence of bound protein will inhibit further crystal growth on this surface.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. S. C. Yang, for his scientific guidance and spiritual support. His provision of an environment for independent growth and exploration is much appreciated. His example as a scientist and a friend also deserves mentioning. Thanks are also extended to the other members of my Research Committee, Dr. G. Wright, Dr. V. S. Ananthanarayanan and Dr. F. Sicheri for their guidance. Many thanks to the lab members Susan Xue, Wai-Ching Hon, Alex Viinberg, Roy Satmaka, Alex Jorov, Victor Pau, and Veronica Yim for their friendship and kindness.

To May, my sisters and brother I give my love for their spiritual support encouragement. Especially May whom stood by me rain or shine. My parents deserve special mention for giving up all their wealth and risked their lives in seek of a place that would provide a better future for their children. I hope that this thesis merit the suffering they had endured.

Last but not least, I would like to extend my gratitude to Canada for opening its arms to a desperate refugee family and providing a home where we can freely explore our potential. Canada is a special place in my

heart and I have made a personal promise to contribute to it with my best

efforts.

Table of contents

List of figures

List of tables

List of abbreviations

ACN   acetonitrile

BGP   bone γ-carboxyglutamic acid protein

BMU   basic multicellular unit

CD   circular dichroism

DLS   dynamic light scattering

FT-IR   fourier-transform infrared

Gla   γ-carboxylated glutamic acid

HPLC   high performance liquid chromatography

MAD   Multiple Anomalous Diffraction

MIR   Multiple Isomorphous Replacement

MW   molecular weight

NMR   nuclear magnetic resonance

PEG   polyethylene glycol

PMSF   phenylmethylsulfonylfluoride

R_h   hydrodynamic radius

RMSD   root mean square deviation

SAS   Single Anomalous Scattering

SDS-PAGE   sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SIR   Single Isomorphous Replacement

TFA   Trifluoroacetic Acid (CF_3COOH)

Chapter 1

General introduction

Bone structure, composition and remodeling

The term bone refers to a family of material that consists of mineralized collagen fibrils (Weiner and Wagner 1998). Bone can be viewed analogous to concrete with collagen as the steel reinforcing bars and mineral crystals act as the cement. The mineral crystals in bone are hydroxyapatite $(Ca_{10}(PO_4)_6(OH)_2)$. These crystals also contain carbonate, magnesium, fluoride, and other impurities (Driessens and Verbeeck 1990); therefore, it may not be precise to call bone crystal hydroxyapatite. However, for the purpose of this thesis, bone crystal will be referred to as hydroxyapatite. The arrangement of collagen fibrils within a collagen bundle varies in different bone types to optimize the corresponding function. There are four most common arrangements of collagen. The arrays of parallel fibrils pattern can be viewed as analogous to a bundle of parallel straws where the straws are collagen fibrils. This structural pattern is found in mineralized tendons, parallel-fibered bone, and fish scales (Ziv, Wagner et al. 1996) reviewed in (Weiner and Wagner 1998). A second arrangement of collagen is the woven fiber structure. This arrangement is more random,

analogous to a loosely packed pile of hay. This type is common in skeletons of amphibia, reptiles, mammalian embryo, and initial bone formed after a fracture where rapid formation is important. A third configuration resembles plywood-like structures. This arrangement is characterized by sets of parallel fibrils present in discrete layers and the fibril orientation in each layer is different, analogous to the arrangement of wood layers in plywood (Weiner, Traub et al. 1999). This type is commonly found in fish scales, cementum that bonds tooth dentin to jaw bone, and lamellar bone that is found in structural bone such as the femur (Weiner, Traub et al. 1999). Finally, collagen can be found in radial fibril arrays. Here the collagen fibrils are almost all in the plane parallel to the surface. Within the plane, the fibril bundles are randomly oriented with respect to each other, but radiates out from the center allowing pores to form, analogous to randomly oriented pile of wood around a well (Weiner and Wagner 1998).

As mentioned above, all bone types consist of mineralized collagen fibrils as their building block. Each fibril is a type I collagen which is made up of three polypeptide chains about 1000 amino acids long. The chains are wound together forming a triple helix. The average diameter of a triple helix is about 1.5 nm and length of 300 nm. In bone, the fibrils are embedded with hydroxyapatite crystals. The bone crystals are plate-shaped with average dimensions of 50 x 25 x 1.5-4.0 nm (Weiner and Wagner 1998).

Because bone is hard and remains intact long after it is dead, many people think that bone is a static material. In fact, bone undergoes constant

turnover like other tissues. The turnover process involves the break down of bone by osteoclasts (specialized cells that break down bone) and simultaneous rebuilding by osteoblasts (specialized cells that lay down new bone). This process occurs at discrete sites named basic multicellular units (BMUs), which contain the activities of both osteoclasts and osteoblasts, though in different regions of the BMU (Compton 2002). During the turnover process, a number of extracellular proteins are produced. Some of the major bone matrix proteins include type I collagen, proteoglycans, bone sialoprotein, bone morphogenic proteins, osteonectin, osteopontin, and osteocalcin. With the exemption of collagen and bone morphogenic proteins, which provide tensile strength and promote differentiation of bone cells respectively, the functions of these proteins are still speculative. Osteocalcin is the subject of this investigation.

Osteocalcin

Osteocalcin is the most abundant non-collagenous protein found associated with the mineralized matrix.

Osteocalcin is a small (46-50 residues) bone specific protein that contains 3 γ-carboxylated glutamic acid residues in its primary structure. The name osteocalcin (osteo, Greek for bone; Calc, Latin for lime salts; in, protein) derives from the protein's ability to bind Ca^{2+} (Hauschka and Carr 1982) and its abundance in bone (Lian, Roufosse et al. 1982; Hauschka, Frenkel et al. 1983;

Conn and Termine 1985). Osteocalcin is also known as "bone γ-carboxyglutamic

acid protein" (BGP) and "vitamin K-dependent protein of bone". It is distinguished

by the presence of 3 γ-carboxylated glutamic acids (Gla), although some human

osteocalcin has been shown to contain only 2 Gla residues (Poser and Price

1979). Because the primary sequence of osteocalcin is highly conserved among

species and it is one of the ten most abundant proteins in the human body

(Gallop, Lian et al. 1980), it is reasonable to infer that the function of osteocalcin

is important.

Osteocalcin structure

The amino acid sequence of osteocalcin has been determined for 22

different species as shown in figure1-1. The primary structure of osteocalcin from

all species share extensive identity, suggesting that its function is preserved

throughout evolution. Conserved features include 3 Gla at positions 17, 21, and

24, a disulfide bridge between Cys23 and Cys29, and most species contain a

hydroxyproline at position 9. The N-terminus of osteocalcin shows highest

sequence variation in comparison to other parts of the molecule. Conformational

study of osteocalcin by circular dichroism (CD) has shown the existence of α-

helical conformation in osteocalcin and that addition of Ca^{2+} induces higher

helical content (Hauschka and Carr 1982; Atkinson, Evans et al. 1995). Two-

dimensional nuclear magnetic resonance (NMR) studies of osteocalcin in

solution, while structurally inconclusive, revealed that the calcium-free protein was effectively unstructured except for the turn required by the disulfide bridge between Cys23 and Cys29. All the proline residues (Hyp9, Pro11, Pro13, Pro15, and Pro27) were in the trans conformation. β-turns are present in the region of Tyr12, Asp14 and Asn26. The hydrophobic core of the molecule is composed of the side chains of Leu2, Leu32, Val36 and Tyr42. The calcium-induced helix is extremely rigid due to, in part, the hydrophobic stabilization of the helical domain by the C-terminal domain (Atkinson, Evans et al. 1995).

```
Monkey    YLYQW LGAPA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Rabbit    QLING QGAPA PYPDP LEPKR EVCEL NPDCD ELADQ VGLQD AYQRF YGPV
Human     YLYQW LGAPA VYPDP LEPRR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Cow       YLDHW LGAPA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Pig       YLDHG LGAPA PYPDP LEPRR EVCEL NPDCD ELADH IGFQE AYRRF YGIA
Sheep     YLDPG LGAPA PYPDP LEPRR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Goat      YLDPG LGAPA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Dog       YLDSG LGAPV PYPDP LEPKR EVCEL NPNCD ELADH IGFQE AYQRF YGPV
Cat       YLAPG LGFPA PYPDP LEPKR EICEL NPDCD ELADH IGFQD AYRRF YGTV
Wallaby   YLYQT LGAPF PYPDP QENKR EVCEL NPDCD ELADH IGFSE AYRRF YGTA
Rat       YLNNG LGAPA PYPDP LEPHR EVCEL NPNCD ELADH IGFQD AYKRI YGTTV
Mouse     YL    GASV  PSPDP LEPTR EQCEL NPACD ELSDQ YGLKT AYKRI YGITI
Xenopus   SYGNN VGQGA AVGSP LESQR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Emu       SFAV  GSSYG AAPDP LEAQR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Chicken HYAQDS GVAGA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Carp      AG    TAPAD LTVAQ LESLK EVCEA NLACE HMMDV SGIIA AYTAY GPIPY
Tetraodon AAGE  PTLQQ LESLR EVCEL NIACD EMADP AGIVA AYAAY YGPPT F
Fugu      APGE  PTPQQ LESLR EVCEL NIACD EMADT AGIVA AYAAY YGPPP F
Bluegill  AAGE  LTLTQ LESLR EVCEA NLACE DMMDA QGIIA AYTAY YGPIP Y
Seabream  AAGQ  LSLTQ LESLR EVCEL NLACE HMMDT EGIIA AYTAY YGPIP Y
Swordfish A TRAGD LTPLQ LESLR EVCEL NVSCD EMADT AGIVA AYIAY YGPIQ F
```

Figure1-1. Amino acid sequence of osteocalcin from 21 different species. The sequences were arranged for maximum alignment. Monkey (Hauschka, Carr et al. 1982); Human (Poser and Price 1980); Cow (Price, Poser et al. 1976); Pig, goat, sheep, and wallaby (Huq, Tseng et al. 1984); Cat (Shimomura, Kanai et al. 1984); rat (Pan and Price 1985); mouse (Celeste, Rosen et al. 1986); rabbit (Virdi, Willis et al. 1991); chicken (Carr, Hauschka et al. 1981); swordfish (Price 1983); Carp, Fugu, Tetraodon (Nishimoto, Waite et al. 2003); seabream (Cancela, Williamson et al. 1995); bluegill (Nishimoto, Arake et al. 1992).

Osteocalcin Activities: Interaction with ionic Ca^{2+} and Mineral Crystals.

Osteocalcin in solution binds Ca^{2+} with a dissociation constant ranging from 0.5 to 3 mM, with a stoichiometry of between 2 and 5 mol Ca^{2+}/mol protein (Hauschka and Carr 1982; Isbell, Du et al. 1993). Osteocalcin binds more strongly to hydroxyapatite (Kd $\approx$ 10^{-7} M)(Poser and Price 1979). It appears that the Gla residues in osteocalcin are important for its affinity toward Ca^{2+} and hydroxyapatite. Binding of Ca^{2+} induces normal osteocalcin to adopt the α-helical conformation; however, thermally decarboxylated osteocalcin showed higher α-helical content than normal osteocalcin and the calcium induced α-helical formation is lost (Hauschka and Carr 1982). Decarboxylated osteocalcin also lost its specific binding to hydroxyapatite (Poser and Price 1979; Hauschka, Lian et al. 1989). When bound to hydroxyapatite, the Gla residues are protected from thermal decarboxylation (Poser and Price 1979). Furthermore, osteocalcin synthesized in animals treated with warfarin, which inhibits the formation of Gla, failed to bind to bone (Hauschka and Reid 1978; Price and Willamson 1981; Pastoureau, Vergnaud et al. 1993). These findings indicate that the Gla residues in osteocalcin are directly involved in Ca^{2+} and hydroxyapatite binding. More recently, fourier-transform infrared (FT-IR) spectroscopic studies shown that the Gla residues in osteocalcin coordinate to Ca^{2+} in the malonate chelation mode,

where a Ca^{2+} interacts with two oxygen atoms, one from each of the two COO-groups of a single Gla residue (Mizuguchi, Fujisawa et al. 2001). It appears that the Gla residues are necessary, but not sufficient for osteocalcin binding to hydroxyapatite, and that the conformation of osteocalcin is also. The binding affinity of osteocalcin for hydroxyapatite increased fivefold by the addition of 5 mM Ca^{2+} (Hauschka and Carr 1982). Furthermore, hydroxyapatite competition studies demonstrated that prothrombin (10 Gla/molecule) and decarboxylated osteocalcin fail to compete with ^{125}I-labeled osteocalcin bound to hydroxyapatite (Hauschka, Lian et al. 1989). These findings suggested that correct protein folding is necessary for osteocalcin's specific hydroxyapatite binding. Combined all the information discussed above, a structural model has been constructed (Hauschka, Lian et al. 1989). This model consists of two antiparallel α-helical domains. The Gla residues are spaced about 5.4 Å apart on one of the helices, which is similar to the interatomic lattice spacing of Ca^{2+} in the x-y plane of hydroxyapatite. It was therefore, predicted that the Gla residues in osteocalcin bind to the 001 plane of hydroxyapatite lattice (Hauschka and Carr 1982; Hauschka, Lian et al. 1989).

Osteocalcin's Biological function

Shortly after the discovery osteocalcin in 1975 (Hauschka, Lian et al. 1975), great efforts were focused on the characterization and functional

determination of this novel protein. In vitro experiments were designed to demonstrate that the Gla residues are important for binding to hydroxyapatite, as described above, and in vivo experiments were designed to study the effect of Gla-lacking-osteocalcin on bone. In addition to osteocalcin's affinity to for hydroxyapatite, it has also been shown that the transition of brushite $(CaHPO_4.2H_2O)$ to hydroxyapatite $(Ca_{10}(PO_4)_6(OH)_2)$ is inhibited by very low concentration of osteocalcin (Hauschka, Lian et al. 1989). This suggests that osteocalcin may function similarly in bone to inhibit or regulate hydroxyapatite formation. The first *in vivo* indication of osteocalcin involvement with the mineralization of bone was demonstrated by Hauschka et. al. that osteocalcin appears in embryonic chick bones coincident with the onset of mineralization (Hauschka and Reid 1978). Studies of bone physiology in animals maintained on warfarin, which inhibits vitamin K-carboxylase, further supports the importance of osteocalcin in bone mineralization. Rats maintained on warfarin during 8 months showed a dramatic closure of the epiphyseal growth plate, causing a cessation of the longitudinal growth (Price, Williamson et al. 1982). Lambs that were maintained on high doses of warfarin from birth to 3 months of age had a significant decrease of trabecular bone turnover, a decrease of bone resorption, and a dramatic reduction of the bone formation rate (Pastoureau, Vergnaud et al. 1993). These animal studies suggest that osteocalcin, possibly along with other Gla-containing proteins, is important for bone turnover. More recently, Ducy et. al. demonstrated that mineralized bone from osteocalcin-deficient mice was two

times thicker than that of wild-type. It was shown that the absence of osteocalcin lead to an increase in bone formation without impairing bone resorption and did not affect mineralization (Ducy, Desbois et al. 1996). Ducy et. al. further suggested that osteocalcin may bind to a specific, yet to be identified, receptor to fulfil its function. As a consequence of this suggestion, Bodine et. al. demonstrated that conditionally immortalized human osteoblasts metabolically responded to osteocalcin in solution. Pretreatment of cells with inhibitors of adenylyl cyclase, phopholipase C, and intracellular calcium release inhibited the response of the cells to osteocalcin. It was concluded that these results indicated that osteoblasts express an osteocalcin receptor, and this putative receptor is coupled to a G-protein (Bodine and Komm 1999). Evidence for the existence of an osteocalcin receptor on osteoclasts has also been demonstrated. Osteocalcin has been shown to induce chemotaxis, cellular differentiation, and calcium-mediated intracellular signaling in osteoclast-like cells, derived from giant cell tumors of bone (Chenu, Colucci et al. 1994). As evident in this summary, information about the functional role of osteocalcin is fragmented and sometimes contradictory, the precise function of osteocalcin and mechanism of action are yet elusive.

Project objectives

Though the precise function of osteocalcin in unknown, it seems clear that osteocalcin binds to bone crystals as well as bone processing cells. Precisely how it binds and the consequences of binding are being actively investigated. The mechanism of osteocalcin's action has been difficult to elucidate due to, in part, the fact that it has no known enzymatic activities. Its activity is apparently conferred only by physical interactions with its target(s), which is undoubtedly dependent on the structural characteristics of osteocalcin. The objective of this project is to determine the crystal structure of osteocalcin and utilize it to gain insight into the function of osteocalcin, specifically, its interaction with hydroxyapatite. Though this project was driven by basic scientific curiosity, it is also hoped that the knowledge gained in this study would assist the development of therapy for bone diseases.

Overview of this thesis

The work presented in thesis initiated as my undergraduate project, where pure bovine osteocalcin was successfully purified. The procedure for purification of osteocalcin is described in Chapter 2. As the term of the undergraduate project ended, I was disappointed that the structure was not determine; therefore, I decided to do volunteer work during the summer to put an end to my "unfinished

work". The following chapters describe this effort. Despite great effort in attempting to crystallize osteocalcin from bovine, I failed to find a suitable condition. It was then decided to expand our search conditions by screening with osteocalcin from a variety of species, instead of focusing on bovine. In addition to bovine, chicken and rabbit osteocalcin failed to crystallize in the conditions tested. Eventually crystals of porcine osteocalcin were obtained. Chapters 3 to 6 describe the procedures and results of structure determination. Each of these chapters is self-contained in that relevant background materials not introduced in this chapter and methods pertinent to the results are discussed. Porcine osteocalcin crystallized as a dimer; therefore, dynamic light scattering was used to determine if such a dimer exist in solution. This work is described in chapter 7. The crystal structure of osteocalcin showed a flat and highly charged surface, which appear to be competent in hydroxyapatite binding. When docked to hydroxyapatite surface, it was clear that osteocalcin can bind with high specificity. This work is described in chapter 8.

Chapter2

Protein preparation and crystallization

Protein Preparation

Materials

Osteocalcin was extracted from its natural source, bone. Bone from a number of different species were purchased from the local supermarkets, slaughter houses, and McMaster central animal facility (CAF).

Methods

Protein production

The production of osteocalcin was carried out based on the previously described protocol (Colombo, Fanti et al. 1993) with modification for scale-up production. The diaphysis of femur bone was separated from the epiphysis with a band saw and flesh was removed from the diaphysis with a razor blade and

wood scraper. After soaking in cold acetone for 10 minutes, the periosteum lining was removed with a wood scraper and steel wool. The marrow was subsequently removed with a long spatula and the medullary cavity was cleaned with a test tube brush in warm soap water. The cleaned bone diaphysis was cut longitudinally in half then crosscut into thin slices (~2 mm) then frozen in liquid nitrogen and lyophilized overnight (FTS Systems Inc.). The lyophilized bone was frozen in liquid nitrogen then ground into powder with a stainless steel blender (Waring Commercial Blender) followed by a coffee grinder (Braun). The bone powder was sieved through a stainless steel mesh yielding bone powder with average size of 200 μm. The fine powder (30 grams) was washed two times with 500 ml of cold water containing 0.6 mM PMSF at 4°C for 30 minutes and the pellet was collected after centrifugation (Sorvall) at 1500 x g for 30 minutes. The pellet was frozen in liquid nitrogen and lyophilized. The freeze-dried powder was demineralized at 4°C by gentle stirring for 4 hours in 300 ml of 20% formic acid (HCOOH) containing 0.6 mM PMSF. The solution was then centrifuged at 40,000 x g for 60 minutes. The insoluble pellet was resuspended with 50 ml of 20% formic acid, stirred for 30 minutes to release any trapped proteins, and recentrifuged. The 40,000 x g supernatants were combined and filtered (Millipore AP2004700).

Protein purification

The filtered supernatant was made to 0.1% Trifluoroacetic Acid

(CF_3COOH) (TFA). Sep-Pak C18 cartridges (Waters No. WAT043345) were

mounted onto a 1-liter flask under vacuum, conditioned with sequential addition

of 100 ml Methanol (CH_3OH) and 100 ml of 0.1% TFA, and then loaded with 100

ml aliquots of the filtered supernatant. After sequential washings with 100 ml of

0.1% TFA and 100 ml of 30% methanol in 0.1% TFA, the bound material was

eluted with 25 ml of 80% methanol in 0.1% TFA into a flask containing 10 □l of

300 mM PMSF. The flow rate was set at 5 ml/minute in the load and elution steps

and 10 ml/minute in other steps. The eluate was concentrated to about 20 ml

with a speed vac (Savant Sped-Vac SC210A). The concentrated solution was

further purified with reverse phase high performance liquid chromatography

(HPLC). The HPLC system consisted of Waters Delta Pak C18-100A 19mm x

30cm column mounted onto Beckman pumps (Beckman 112 Solvent Delivery

Module), which are controlled by a BioRad Chromotograph software on Windows

95. The elution gradient was created by mixture of two eluants, eluant A

consisted of HPLC grade acetonitrile (ACN) and eluant B consisted of 0.1% TFA.

The gradient profile was set from 25% to 45% ACN over 30 minutes. The flow

rate was set at 10 ml/minute. The well-resolved peak that eluted after 40 minutes

(39% ACN) (see figure 2-1) was collected, frozen in liquid nitrogen and

lyophilized overnight. The freeze-dried protein was stored at -20 °C.

Results

Typically, about 10 mg of protein was purified from 30 grams of bone powder. The purity of the isolated protein was assessed with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and analytical HPLC. On SDS-PAGE, only one band can be seen (figure 2-2) and on analytical HPLC, only one peak is apparent (figure 2-3). The authenticity of this purified protein was assessed with mass spectrometry (figure 2-4) and amino acid composition. The mass of the predominant peak in the mass spectrometry spectrum (peak A) was in good agreement with the calculated molecular weight (MW) for osteocalcin (calculated 5737.19 Dalton and measured 5737.21), and amino acid composition agreed well with the composition of osteocalcin.

Figure 2-1. Preparative reverse phase HPLC of porcine osteocalcin. Waters delta-pak

C_{18}-100A 19 mm x 30 cm column equilibrated with 25% CH_3CN in 0.1% TFA and eluted

over 40 minutes with a linear 25-45% CH_3CN gradient in 0.1% TFA. Flow rate, 10ml/min.

Solid line, absorbance; broken line, CH_3CN gradient.
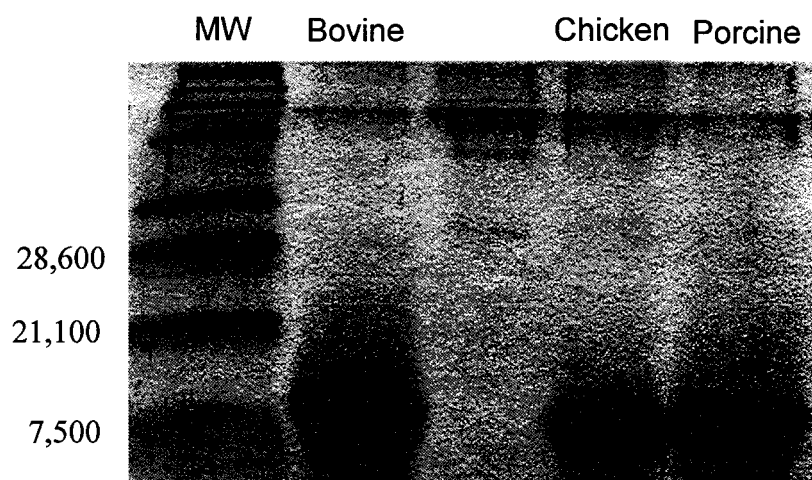
Figure 2-2. SDS-PAGE of purified bovine, rabbit, chicken, and porcine

osteocalcin. HPLC purified osteocalcin from bovine (lane 2), chicken (lane 4), and

porcine (lane 5) were loaded onto 15% SDS-polyacrylamide gel with 4% stacking along

with BioRad Broad Range Molecular Weight marker.

Bio-Rad Laboratories ValueChrom System
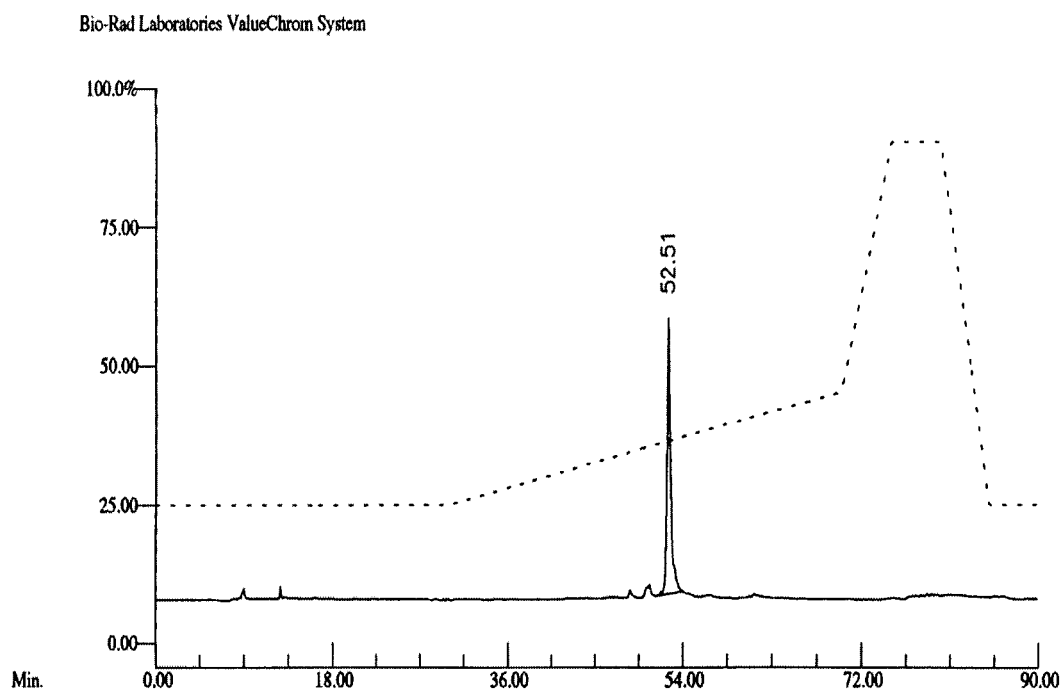
Figure 2-3. Reverse phase HPLC of purified porcine osteocalcin. Waters delta-pak C_{18}-100A 3.9 x 300 mm column equilibrated with 25% CH_3CN in 0.1% TFA and eluted over 50 minutes with a linear 25-45% CH_3CN gradient in 0.1% TFA. Flow rate, 1ml/min. Solid line, absorbance; broken line, CH_3CN gradient.
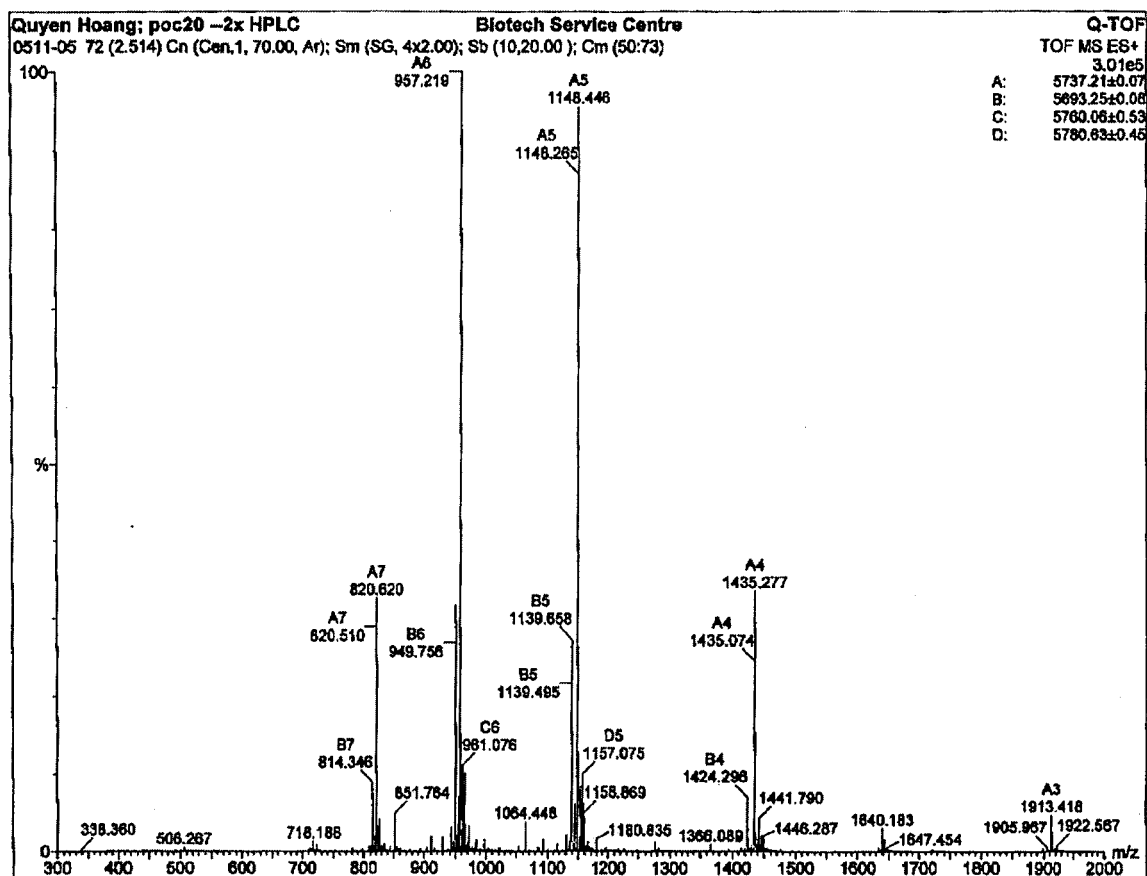
Figure 2-4. Mass-spectrometry of purified porcine osteocalcin. Peak A has an apparent molecular weight of 5737.21 Da, which corresponds to calculated weight of porcine osteocalcin.

Crystallization

Materials

HEPES and calcium chloride were purchased from Sigma. PEG 4000 was purchased from Fluka. Crystallization trays and crystal screening kits were purchased from Hampton Research. Circular cover slips used for vapor diffusion crystallization were purchased from VWR. All solutions used for crystallization were pre-filtered through 0.2 □m membrane filtration units (Millipore).

Methods

Crystallization was performed by the vapor diffusion method (McPherson 1989) in hanging drop mode for crystal screening and in sitting drop mode for diffraction quality crystals. Circular cover slips (22 mm) were silanized by immersion in dimethyldichlorosilane for 5 minutes, rinsed liberally with water then baked at 190 °C for 1 hour. For hanging drops, a small bead of grease was placed on the rim of each well in the crystallization tray. Typically 500 ml of mother liquor was placed into the well. The purified protein was dissolved in solution to 10 mg/ml and 2 µl of the protein solution was mixed with 2 µl of mother liquor on the silanized cover slip with or without the addition of calcium chloride. The cover slip was then sealed over the well in the inverted position, such that the mother liquor and the drop on the cover slip share the same air

space. Two identical sets of screens (196 conditions each) were set up; one set was allowed to equilibrate at room temperature and the other at 4 °C.

Results

Screens were set up for osteocalcin from bovine, chicken, rabbit and porcine, only samples from porcine bone produced crystals. After the crystal screens were allowed to equilibrate for 1 week, 27 conditions with added calcium chloride produced crystals of various morphologies and qualities at room temperature. Conditions of Hampton Research Screen I, which consisted of 10% v/v isopropanol, 0.1M HEPES-Na pH 7.5 and 20% w/v PEG 4000, were chosen for further fine-tuning. This mixture was chosen because it consisted of no salt, at physiological pH, and contains PEG 4000, which would alleviate the complication of searching for a cryo-protectant. The mixture was further refined with the objective of producing the best crystal quality while reducing the amount of isopropanol to reducing complications with crystal handling. The final refined mixture consisted of 0.1M HEPES pH 7.5 in 10% w/v PEG 4000. This condition produced large diffraction quality crystals (see figure 2-5).
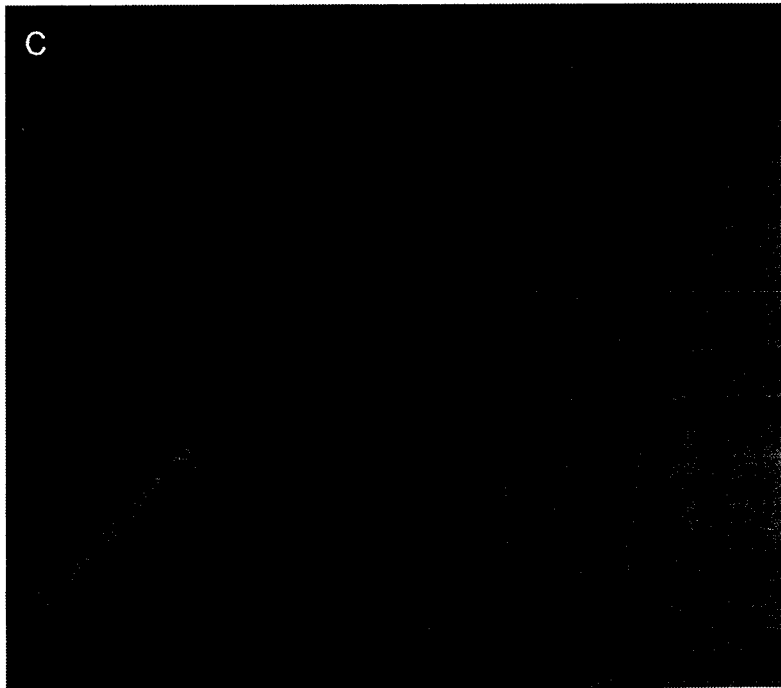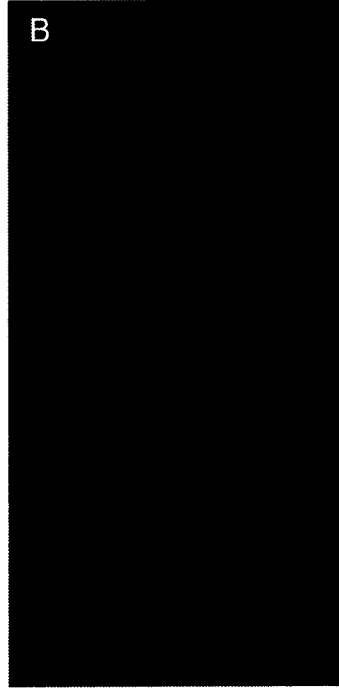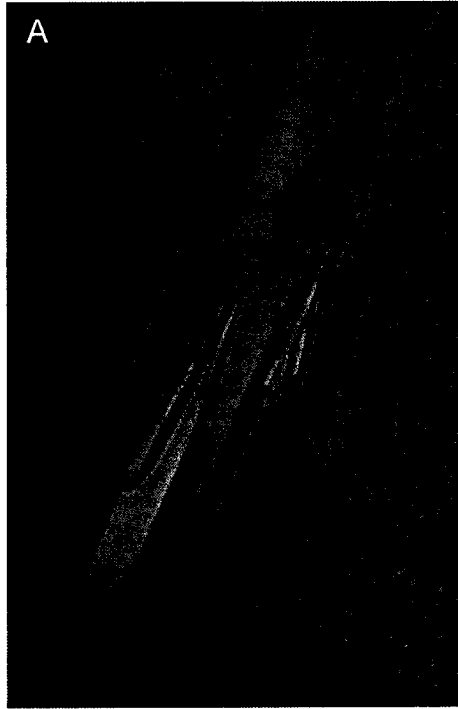
Figure 2-5. Large-scale crystallization of porcine osteocalcin using sitting drop method.

Purified protein was dissolved in 100mM HEPES pH 7.5 to a concentration of about

10mg/ml. Mother liquor solution contains 100mM HEPES, 10% w/v PEG 4000 and

10mM $CaCl_2$. The sitting drop contains 20μl of protein solution and equal volume

of mother liquor. The size and number of crystals in a drop can vary in different

batch. The dimension of the crystal in A was about 200 x 600 μm; B, 100 x 300

μm; C, 50 x 400 μm.

Chapter3

Data Collection and Processing

Introduction

Many people feel that x-ray crystallography is abstract and the processes involved are commonly referred to as a "black box". In reality, it is as real and no more of a black box than the magnification process that occurs in a common light microscope. In the light microscope, visible light is scattered from the specimen. The diffracted waves are bent by the lenses of the microscope, in ways that the relative phases of the waves are maintained, to provide the magnification of the specimen and focusing of the magnified image. In x-ray crystallography, the specimen (the protein crystal in this case) scatters x-radiation similar to the diffraction of visible radiation in the light microscope; however, in place of the objective lens, the diffracted waves are intercepted with a detector (x-ray films, CCD chips, phosphor imaging plates, etc.). The detected image is a pattern of spots that represents the reciprocal lattice. Each of these spots contains information of the whole unit cell. Unlike the visible radiation, there is currently no means to focus the diffracted x-ray waves; therefore, the image of the protein cannot be visualized. The relationship between the sample and its diffraction

pattern is physical; therefore, the image of the sample can be calculated from its diffracted images. To calculate the molecular arrangements of the protein molecules, the position, intensity, and relative phase of each diffracted spot must be known. The positions and intensities of the spots are easily measured; however, the relative phases of the waves are lost upon impacting the detector. This missing part of the equation makes the reconstruction of the protein molecule (the diffracting specimen) more challenging; this is the so-called "phase problem".

With today's technology, the phase problem cannot be solved directly. There is no way to bend x-radiation significantly to focus the diffracted rays. Without solving the phase problem, methods have been developed to work around the problem. The idea is to estimate what the lost phases may have been, a model is calculated based on the estimated phases and the model is refined to fit the observed data. When the structure factors calculated from the model are in good agreement with the observed diffraction data, the model is said to represent the protein molecule in question.

Diffraction from crystals

A crystal is defined by regular repetitive arrangements of atoms or molecules. Each repeating unit is called a unit cell. The unit cell is chosen to be the smallest repeating unit with identical environments at the corners of the cell.

Due to the ordered arrangements of the atoms in the unit cell, infinite number of planes can be drawn to intersect each unit edges. The diffraction spots observed on the detector are the x-rays that "reflected" from these planes. These planes are identified by a unique index. The index is defined by the reciprocal of the fractions of the unit cell lengths. For example, the reflection from a plane that bisects the edges of a unit cell with the dimensions a, b, c; the fractional distances between the intersecting planes along the cell axes are a/2, b/2, and c/2, the index (h, k, l) is 2 2 2. Each diffracted spot must obey Bragg's law, which states that diffraction occur only in conditions where the path difference between the waves are multiple of the wavelength, $n\lambda=2d\sin\theta$; where n is an integer, λ is the wavelength, d is the distance between the diffracting planes, and θ is the angle between the incident radiation and the diffracting plane. Using these relationships, one can determine the unit cell dimensions by analyzing the diffraction pattern.

Data collection strategies

The objective was to solve the structure of osteocalcin by obtaining the phasing information from native anomalous scatterers, by a method that is called Single Anomalous Scattering (SAS). The theoretical foundation for this method is discussed in chapter 4. The native anomalous scatterers with usable signals in osteocalcin are the bound calcium ions and the two sulfur atoms on the cysteine

residues. The anomalous signals from these elements are very small; the imaginary components of the anomalously scattered waves ($\Delta f''$) for calcium and sulfur are 1.4 and 0.6 electrons, respectively, when the incident X-radiant wave is 1.54 Angstrom in length. Since the signal is so small, the errors in the data must be minimized in order to detect it. The strategies used to minimize the errors are high redundancy, fine slicing, and inverse beam geometry. Assuming that the errors or noise in the data are random, the noise would cancel each other when the same diffraction spot is sampled many times (high redundancy). The intensities recorded from the detector are the projections of the spherical diffraction spots; therefore, sampling the sphere in many different slices would give a more accurate estimation of its intensity. The crystals used for the diffraction experiments are not spherical; therefore, the thickness of the x-ray path through the crystal differs as the crystal rotates, thus error is introduced due to the difference in absorption coefficient. To minimize this effect, only the reflections that are separated by 180 degrees are merged (inverse beam). The size of the signals is a function of the incident x-ray wavelengths; therefore, data was also collected at various wavelengths (1.0 Å, 1.54 Å, and 1.7 Å). Over 50 data sets were collected; however, only 12 data sets were analyzed beyond the data processing stage. These data were selected based on their mosaicity, and symmetry merging R-factors.

Data Collection

Initially, attempts were made to solve the structure of osteocalcin using copper radiation (1.54 Å). In this attempt, five high-resolution data sets (2.4 to 2.0 Angstrom) that produced consistent anomalous Patterson peaks were obtained. Two of the five data sets were collected on the Rigaku Raxis II imaging plates (home source), two sets were collected on the Mar Research Mar345 imaging plates in Toronto, and one was collected on the Bruker SMART 6000 4K CCD detector in Madison, WI. The reflection images collected on the Rigaku and Mar machines were indexed using the program Denzo, scale and merged with Scalepack. The images collected on the Bruker were analyzed by the Bruker SMART program package. All data sets gave the same crystal system and unit cell dimensions. A considerable amount of time and effort was put into trying to solve the structure with these 1.54 Å data sets. However, the resulting electrondensity maps were not interpretable, presumably due the fact that the signals were too weak. Attempts were made to increase the signal by performing the diffraction experiment with radiations of longer wavelengths. Eighteen data sets were collected at the Advanced Photon Source in Chicago. Only 5 data sets were analyzed beyond data processing, one set at λ=1.0 Å, and four sets at λ=1.7 Å. The data was processed with the programs Denzo and Scalepack. The unit cell parameters were virtually identical to the data collect with 1.54 Å radiation; the resolution was a little higher with the synchrotron data; the merging

R-factor was 4.9%. Only the data collected at 1.7 Å wavelength had enough anomalous signals for their location. Due to the higher anomalous signal at this wavelength, no special data collection strategy was applied. The experimental details are as follow; single anomalous scattering (SAS) data were collected at beamline IMCA-CAT ID-17 of the Advanced Photon Source at Argonne National Laboratory. The x-ray wavelength was set at 1.7 Å to maximize the anomalous signals. Osteocalcin crystals were flash frozen in a nitrogen cold stream after transferred to a cryo-protectant solution containing 30% PEG 4000, 10 mM $CaCl_2$, and 0.1 M HEPES at pH 7.4. The x-ray detector, Mar Research m165 CCD, was placed 50 mm from the crystal. Data was collected in oscillation mode from 0 to 180 degree and diffraction images were recorded for each degree of rotation at 3.0 sec per frame (a sample is shown in figure 3-1). The crystal was kept frozen at -160 °C during data collection using the Oxford cryosystem cooling device. The intensities of the diffraction data were integrated and indexed using DENZO and reduced using SCALEPACK of the HKL package (Otwinowski and Minor 1997). The Bijvoet reflections were scaled separately. The statistics of data collection and processing are presented in table 3-1.

Osn33 Frame #30

2.0A →
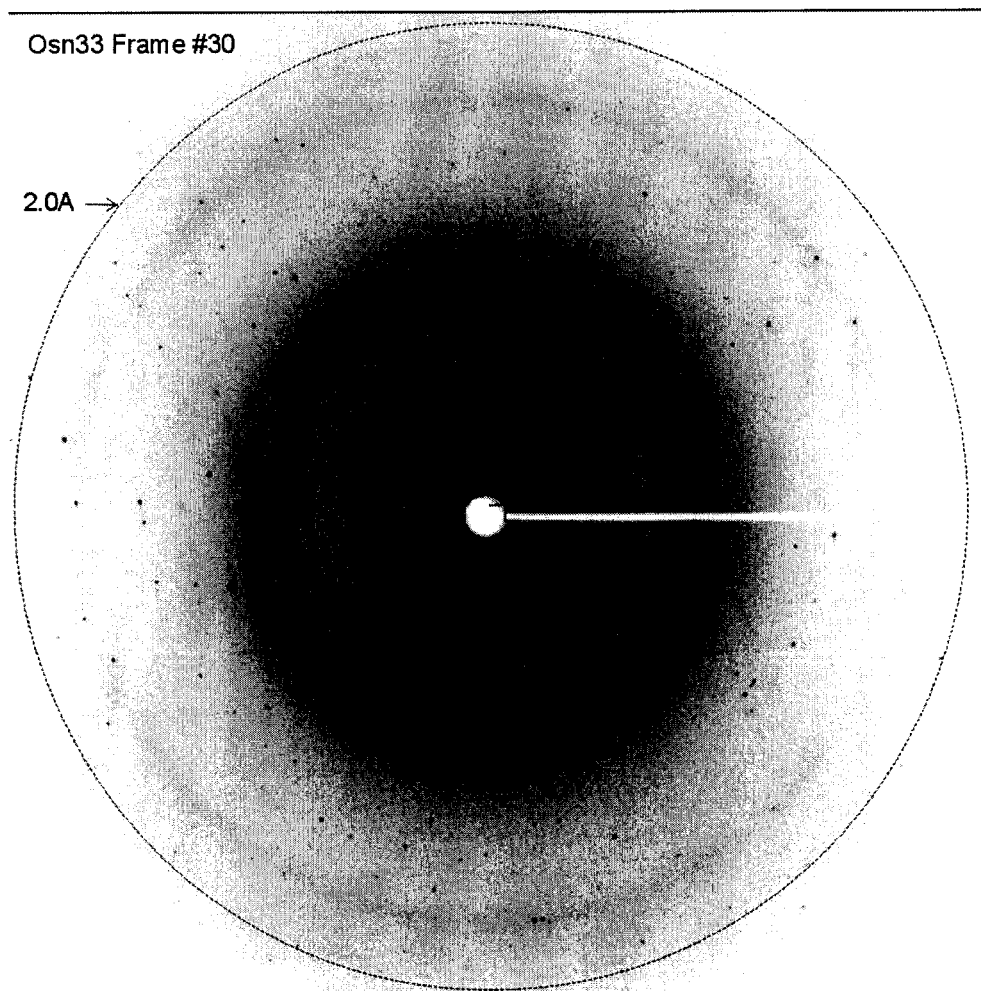
Figure 3-1. Diffraction pattern of porcine osteocalcin crystal. Diffraction image

was collected with a Mar Research m165 CCD detector at IMCA-CAT beamline

at APS. The dotted circle indicates a resolution shell of 2.0 Å. The optical

resolution of this data was 1.8 Å.

Table 3-1. Statistics of data collection and processing of osteocalcin. A, observation redundancy; B, reflection intensities; C, R-merge. Data generated with the program Denzo and Scalepack. R linear = $\Sigma\,|I - <I>|\,/\,\Sigma\,(I)$; R square = $\Sigma\,(I - <I>)^2\,/\,\Sigma\,(I^2)$; Chi^2 = $\Sigma\,(I - <I>)^2\,/\,(Error^2 \times N\,/\,(N-1)\,)$

A. Summary of observation redundancies by shells

Shell Upper Lower Limit Limit	No. of reflections with given No. of observations							
	0	1	2	3	4	5-6	7-8	Total
99.00-4.31	16	20	78	364	246	1	0	709
4.31-3.42	6	13	47	424	212	3	0	699
3.42-2.99	6	13	46	430	209	6	0	704
2.99-2.71	12	16	50	405	208	9	0	688
2.71-2.52	15	22	62	384	230	7	0	705
2.52-2.37	21	22	77	347	245	3	0	694
2.37-2.25	33	21	97	297	258	6	0	679
2.25-2.15	34	24	112	264	258	2	0	660
2.15-2.07	28	25	124	234	291	6	0	680
2.07-2.00	42	24	129	234	300	2	0	689
All hkl	**213**	**200**	**822**	**3383**	**2457**	**45**	**0**	**6907**

B. I/Sigma in resolution shells

Shell	No. of reflections with I/σ less than								
Upper Lower Limit Limit	0	1	2	3	4	10	20	>20	Total
99.00-4.31	0	1	3	5	14	28	80	629	709
4.31-3.42	1	1	1	2	4	18	81	618	699
3.42-2.99	0	5	7	21	39	70	207	497	704
2.99-2.71	2	9	19	32	50	103	368	320	688
2.71-2.52	6	26	46	61	101	222	561	144	705
2.52-2.37	4	24	46	71	135	251	573	121	694
2.37-2.25	13	36	79	113	170	312	594	85	679
2.25-2.15	16	46	101	149	221	383	584	76	660
2.15-2.07	14	81	165	224	333	480	631	49	680
2.07-2.00	33	119	205	284	420	590	671	18	689
All hkl	89	348	672	962	1487	2457	4350	2557	6907

C. Summary of reflection intensities and R-factors by shells

Shell Å	Average I	Chi^2	Linear R-factor	Square R-factor
99.00-4.31	17054.9	1.011	0.035	0.044
4.31-3.42	12599.8	0.924	0.039	0.047
3.42-2.99	6881.3	0.855	0.044	0.053
2.99-2.71	3458.2	0.935	0.055	0.064
2.71-2.52	1853.4	0.884	0.072	0.080
2.52-2.37	1549.1	0.997	0.080	0.088
2.37-2.25	1183.1	1.033	0.099	0.137
2.25-2.15	848.0	1.212	0.128	0.198
2.15-2.07	626.9	1.307	0.157	0.373
2.07-2.00	426.1	1.226	0.214	0.542
All reflns	4718.0	1.036	0.049	0.050

Chapter 4

Phase determination

Introduction

The methods commonly used to obtain phasing information for the initial model are Single Isomorphous Replacement (SIR), Multiple Isomorphous Replacement (MIR), Multiple Anomalous Diffraction (MAD) and less commonly used Single Anomalous Scattering (SAS).

The so-called anomalous scattering refers to the scattering of X rays when the vibrations of the electrons that are excited by the incident X-ray are in resonance with the natural frequencies of vibration of the bound electrons. This resonance causes a slight delay in scattering, which results in a change in intensity of the diffracted rays. The structure factor for an atom with anomalous scattering is more complicated than that of an atom with only normal scattering because the anomalous effect must be accounted; hence, $f = f^o + \Delta f' + i\Delta f''$ where f^o is the normal scattering factor, $\Delta f'$ is the real component of the anomalous scattering and $\Delta f''$ is the imaginary of anomalous scattering. Since the imaginary component is always positive and $\pi/2$ out of phase of the real component (90

degrees ahead of $\Delta f'$), the intensities of the Friedel mates (reflections that are separated by π or 180 degrees) are slightly different. The usefulness of anomalous scattering was first recognized by Bijvoet half a century ago (Bijvoet 1949). He suggested that it could be used to determine the absolute configuration of handed molecules and to resolve phase ambiguity in the single isomorphous replacement experiments (SIRAS). Indeed, anomalous signals are routinely used for these purposes and had since been extended to include direct phase determination (Bijvoet 1954; Hendrickson, Smith et al. 1985), such as MAD and SAS.

The Patterson method is commonly used for the location of anomalous scatterers, where the difference between the Friedel mates or Bijvoet pairs ($\Delta F = |F^+| - |F^-|$) are used as the coefficient in the Patterson function, $P(u,v,w) = 1/V\sum\sum\sum|\Delta F|^2 \exp[-2\pi i(hu + kv + lw)]$. The Patterson method consists of evaluating a Fourier series for which only the amplitudes and the indexes of the diffracted beam are required. The map of this function is a three dimensional grid that fills a space with the same size and shape of the unit cell. The peaks in this map represent the inter-atomic vectors of the atoms in the unit cell. The height of each peak is proportional to the multiplication of the atomic number of the two corresponding atoms. In a protein crystal, usually only the heavy atoms produce peaks that are distinct from the background. The position (x, y, z) of the heavy atoms can be determined from the Patterson map by considering all the symmetry related positions and the positions of Patterson peaks (u, v, w). For

example, a single atom at position (x, y, z) in the space $P\bar{1}$ would have a

symmetry related atom at (-x,-y,-z). The inter-atomic vector between the two

atoms would be at (2x, 2y, 2z); therefore the values of (x, y, z) = (u/2, v/2, z/2).

Once the positions of the heavy atoms are known, the phase of each reflection

can be calculated. In vector representation, the structure factor can be

represented as $|F(hkl)|^2 = [A(hkl)]^2 + [B(hkl)]^2$, where $A(hkl) = \sum f \cos 2\pi(hx + ky$

$+ lz)$ and $B(hkl) = \sum f \sin 2\pi(hx + ky + lz)$; therefore, the phase of F(hkl) can be

calculated with the relationship $\alpha = \tan^{-1} [B(hkl)/A(hkl)]$. The phases of the heavy

atoms (phases of the partial structure) are than extended for all atoms in the

crystal, and from which an electron-density map is calculated.

An inherent problem with the phases determined with the SAS method is

that each of the phases determined is a bimodal distribution where either

distribution has an equal probability of being correct, but only one is the correct

solution. This is referred to as the biphasic nature of SAS, which leads to phase

ambiguity.

In 1985, Bi-Cheng Wang described a procedure that he has developed to

resolve the phase ambiguity inherent with SIR and SAS data; he called the

process *filtering of errors* (Wang 1985). This process is also commonly referred

to as *solvent flattening* and *density modification*. The conceptual foundation of

this approach is that an SIR or SAS electron-density map is a superposition of

two Fourier maps, one being a correctly phased protein density map, and the

other being a map produced with false phases. The map produced with the false

phases will contain no structural information and will be in a form of general background. Its densities are random noise and should be randomly distributed throughout the map. On the other hand, the densities arose from the correct phases would be higher in the protein regions and lower in the solvent regions. This means that the high densities in the solvent regions are mainly due to noise, and when they are removed, a large portion of the noise is filter out. The resulting map contains phases are now biased toward the correct ones. A number of different software programs have been developed based this original idea. The mathematics in determining the protein and solvent boundary and the choice of tasks in each spaces (direct space and reciprocal space) are slightly different in different programs.

Methods

Location of anomalous scatters and phase extension

The positions of calcium and sulfur atoms were located using Bijvoet difference Patterson functions. Automated Patterson search as well as phase determination were performed with the program SOLVE (Terwilliger and Berendzen 1999).

Resolving phase ambiguity

Solvent flattening was carried out with the program RESOLVE (Terwilliger 1999).

Results

Automated Patterson search with SOLVE found 5 peaks which corresponded to 3 calcium and 2 sulfur atoms (see table 4-1).
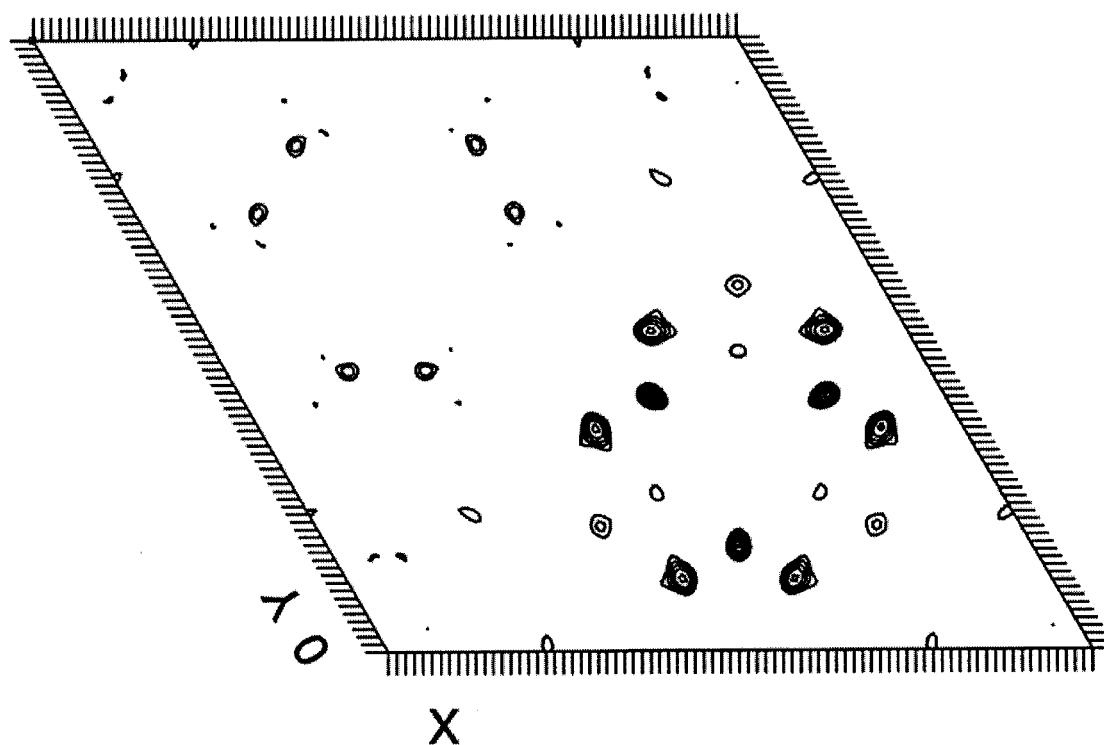
Figure 4-1. Harker section (x,y = 0; z = 0.666) of difference Patterson map.

Patterson map was calculated with SAS $(\Delta F_{\pm h})^2$ as Fourier coefficients. Map is

contoured at 0.5σ interval, starting from 2σ level. Map was generated with the

program CNS.

Table 4-1. Sites of anomalous scatters found by the program SOLVE.

Top Solutions Found by SOLVE					
X	Y	Z	Occup	B	Hight/Sigma
0.396	0.455	0.085	0.296	26.7	13.1
0.418	0.418	0.997	0.166	25.5	14.9
0.576	0.054	0.148	0.347	30.6	12.7
0.366	0.941	0.162	0.198	54.7	5.4
0.476	0.881	0.037	0.250	46.7	6.7
Final Resolution: 2.0 Å					

Chapter 5

Model building and refinement

Introduction

After the phases are calculated, an electron-density map can be calculated by Fourier synthesis; $\rho(x, y, z) = 1/V \ \Sigma_h\Sigma_k\Sigma_l \ F_{hkl} \ e^{-2\pi i(hx + ky + lz)}$. This is the experimental electron-density map into which the initial model is built. The density is displayed electronically on a computer monitor as wire cages. It is more convenient to interpret the map when only part of the density is displayed, ideally containing the whole of a single molecule. It is common practice to display only the asymmetric unit initially, however, in high symmetry space groups a larger portion is needed. The model building process involves fitting the protein molecule in to the density. First, an α-carbon chain is fitted into a continuous density that resembles a polypeptide chain or secondary structures. The direction of the chain is then determined by orienting a helix vertically with the side-chains projecting downwards, analogous to a Christmas tree. In this orientation, the N-terminal would be at the bottom of the tree and the C-terminal is at the tip. Side chains are then fitted after the identification of a few prominent side-chains, such as phenylalanine and tyrosine, which are used as the reference points with

respect to the rest of the primary sequence. This is trial model, which is expected to contain errors and needs to be further refined.

The purpose of refinement is to make the model agree better with the observed diffraction data. The first round of refinement involves refining only the position and orientation of the molecule. Molecular parameters such as bond lengths and angles are fixed, this type of refinement is referred to as rigid body refinement. After rigid body refinement, positions of all atoms in the model are refined, allowing all movements independent of each other within the specified restraints (bond lengths, bond angles, and van de Waals contact distances) and constraints (non-crystallographic symmetry and restriction to torsion angle movement). An electron-density map is then calculated based on the refined model and is used for model rebuilding. The re-building and refinement processes are iterated until the difference map (Fo - Fc) is flattened, or at least the peaks are accounted for. The final step of structure determination is structure validation. There is no way to tell if the final model accurate represents the actual distribution of atoms in the crystal. However, the final model must make chemical sense, i.e., protein folding geometry, bond lengths, bond angles, and non-covalent distances must agree with expected values within the experimental errors. If the model were wrong some anomalies in the model would exist.

Methods

The program O (Jones, Zou et al. 1991) was used for model building and

rebuilding. Model building was initiated using the 2.0 Å electron density map

calculated with SAS phases that were improved by solvent flattening. Because of

the intrinsic handed ambiguity of the space group $P3_121$ and $P3_221$ (they give

rise to identical diffraction patterns) and the handed ambiguity of the SAS peaks,

four maps were generated. Of the four, space group $P3_121$ with positive peaks

had the best phasing statistics and yielded interpretable electron-density map;

therefore, it was used for model building. Though the experimental density map

was excellent, initial tracing was complicated by the tight packing around the β-

turn and close interaction between Gla (residues 17, 21 and 24) and calcium

ions, causing non-covalent density continuity in these regions. The backbone

was fitted in the density using a dipeptide baton provided in the program O. The

chain of dipeptide was then converted to an α-carbon trace of a poly-alanine

chain and the peptide bonds were regularized. Tyrosine 42 was the most

prominent feature in the density and was therefore chosen as a point from which

the sequence was extended. In reference to Tyr42, the alanine chain was

mutated to reflect the amino acid sequence of porcine osteocalcin with the side-

chains placed in the most common conformation. Because Gla is not a standard

amino acid and its parameters were not included in O, the O library was modified

to include the necessary parameters. The Gla side-chains were build into the

osteocalcin sequence manually by duplication of the glutamic acid side-chains and edited by a text editor Jot. The side-chains that extended outside of the density cage were rotated to best fit the density before refinement. Refinement was carried out with the program CNS (version 1.11) (Brunger, Adams et al. 1998). Positional and simulated annealing refinement with maximum likelihood targets was carried out after rigid body refinement. Iterative cycles of model building and refinement were carried out until the x-ray residual factor, R-factor, was stationary and no more information can be obtained from σ_A-weighted (Read 1986) $2|Fo| - |Fc|$ electron-density map, difference map (Fo - Fc), and omit map.

Results

Electron-density and refined model

The statistics of refinement are provided in table 5-1. The refined model along with the experiment electron-density map is shown in figure 5-1. As shown in figure 5-1, the density only covered residues 13 to 49; therefore, only those residues were modeled. Much effort was put towards resolving the first 12 residues without success. The lack of success was probably due to the lack of order in the un-modeled regions.
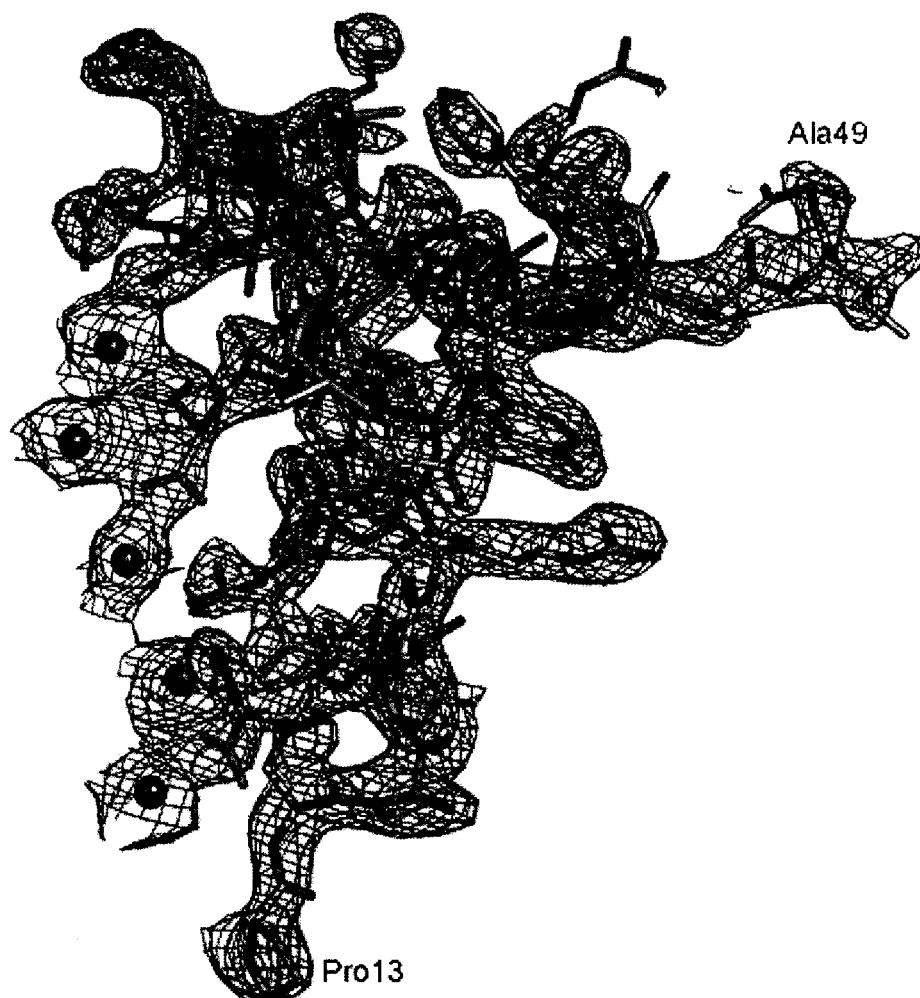
Figure 5-1. Crystal structure of porcine osteocalcin. a, Experimental electron density map for $POC_{13\text{-}49}$. $POC_{13\text{-}49}$ is shown as stick model, where carbon, oxygen, nitrogen and sulfur atoms are gray, red, blue, and orange, respectively. The solvent-flattened SAS map (contoured at 1.5σ) is shown as green mesh. Calcium ions are shown as purple spheres. This image was generated by Pymol.

A

Arg19

Pro18

B

Tyr42
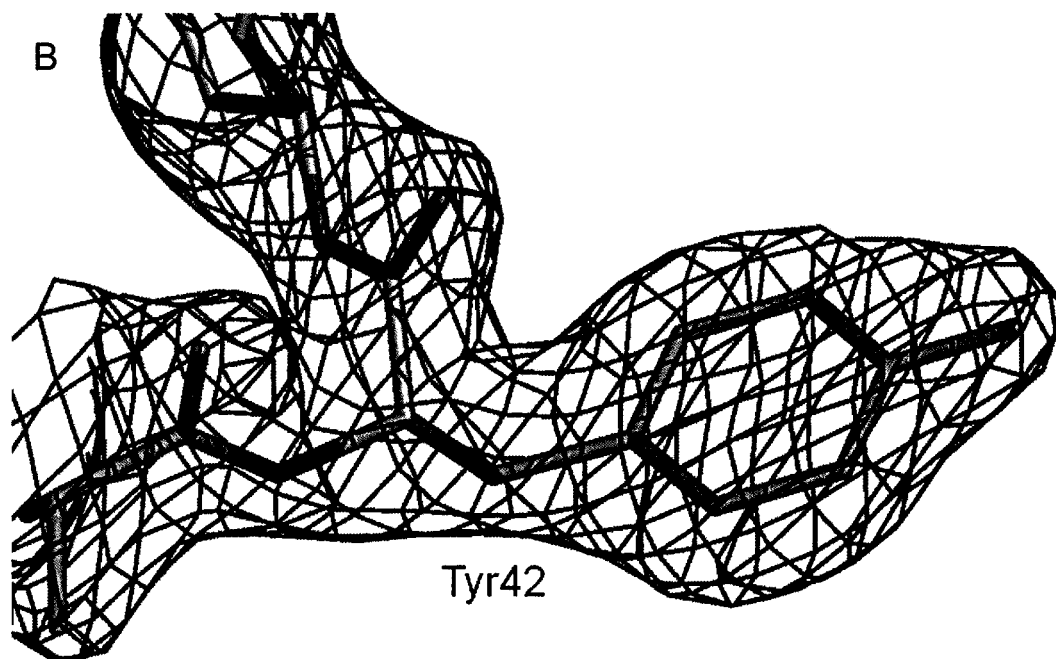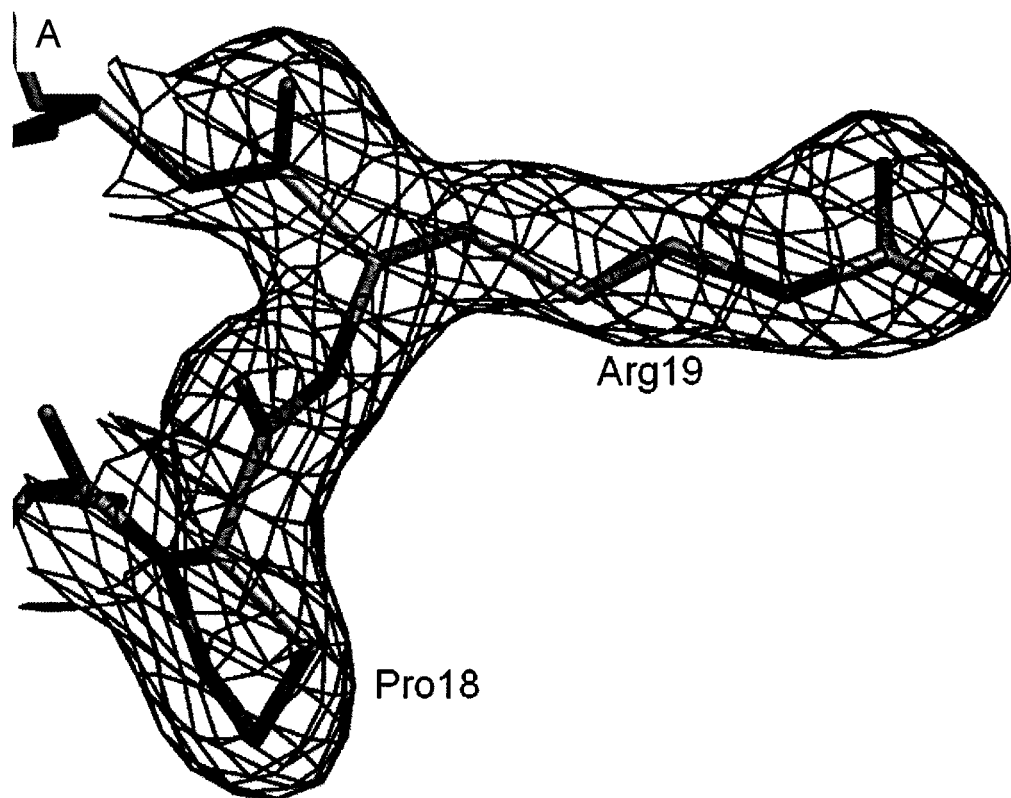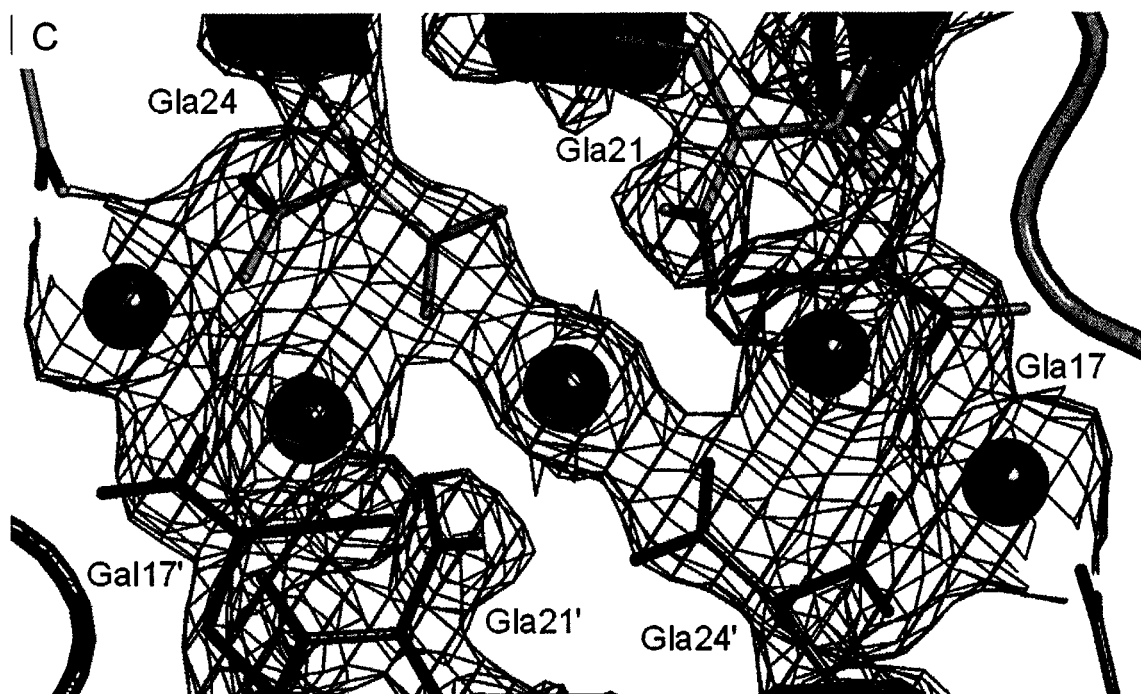
Figure 5-2. Experimental electron density surrounding Arginine 19 and Tyrosine 42 and the bound calcium atoms. A and B, the molecules are shown as stick models, where carbon, oxygen, and nitrogen atoms are gray, red, and blue, respectively. The solvent-flattened SAS map (contoured at 1.5σ) is shown as aqua mesh. C, one monomer is coloured orange and the other blue. Calcium is shown as purple spheres and density shown as marine mesh.

Structure validation

The R-factors for the refined model were reasonably good; the R-factors for most protein structures are between 20 and 30 percent. However, for such a small protein, the R-factors are higher than what was expected. This was likely due to some disorder in the crystal. Such disorder may be consistent with the unusual fall-off of the diffraction pattern, where the diffraction intensities abruptly decreased beyond 2.0 Å resolution. Anyhow, the model matched well with the experimental electrondensity map, RMSD of bond lengths and angles are good (table 5-1), and the phi and psi angles of residues are within the allowed regions (figure 5-3). Therefore, it is reasonable to conclude that the model closely represents the true structure.

A

PROCHECK

Ramachandran Plot
Osteocalcin

B

| Sequence & Ramachandran regions | Most favoured ▲ | Allowed ▮ | Generous▮ | Disallowed |

AAAAAAAAAAAAAAAAAAAA▮A AAAAAAA▮ A
PDPLCPRRCVCCLNPDCDELADHIGFQEAYRRFYGIA
     g   g   g
     u   u   u
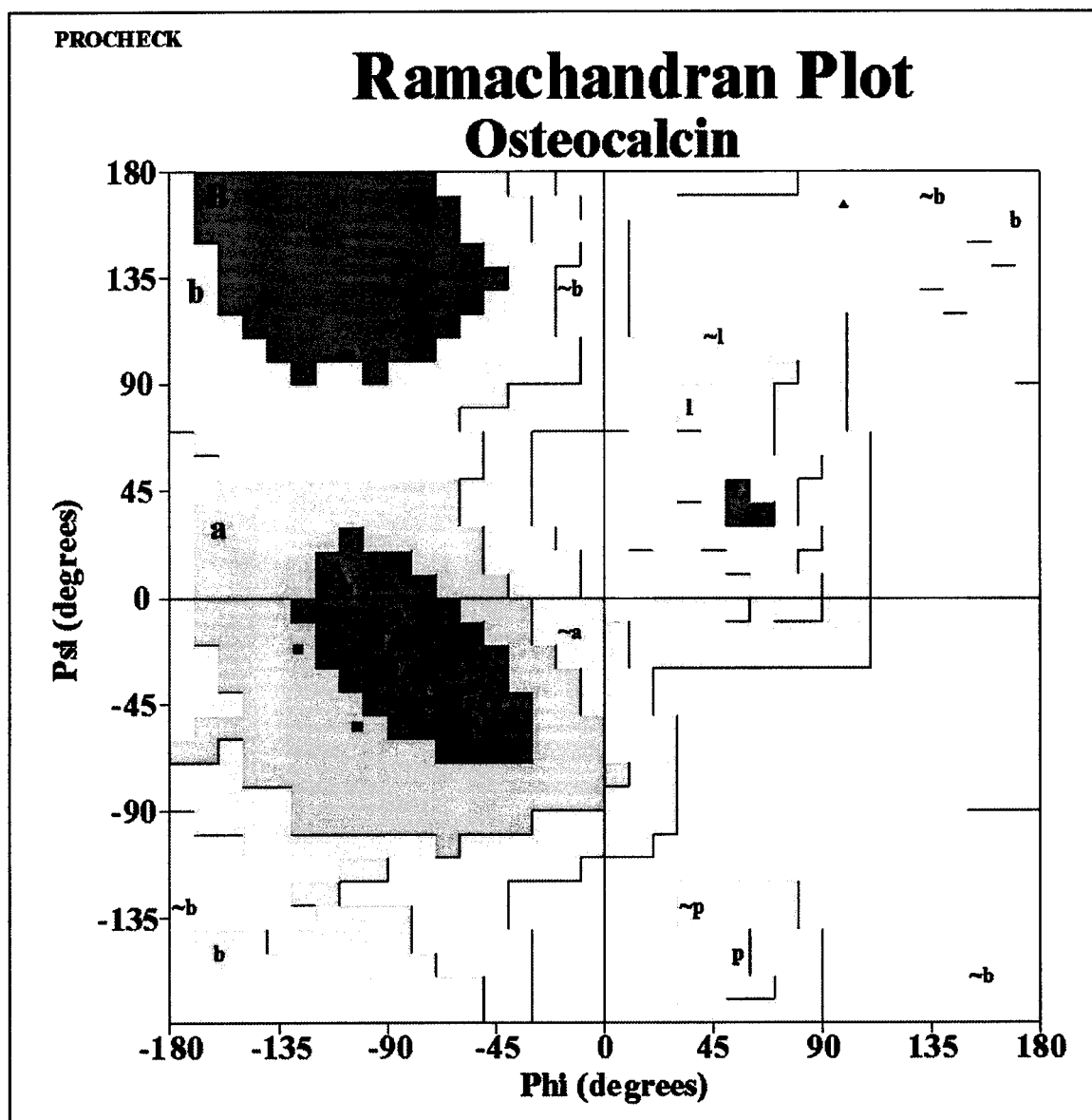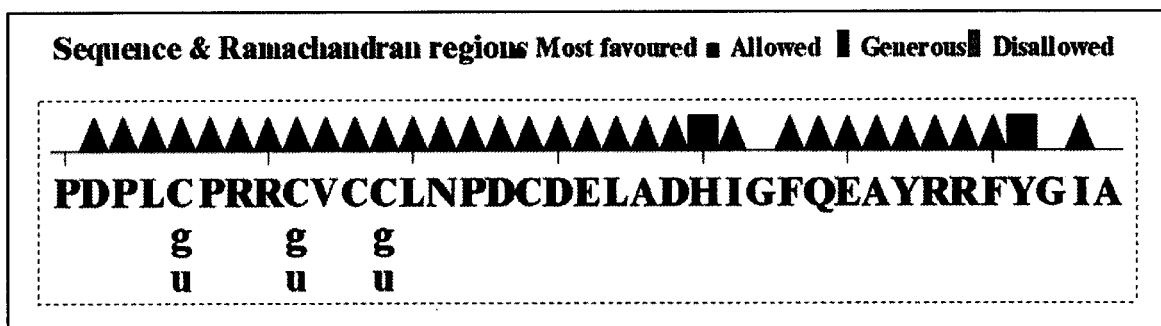
Figure 5-3. Main chain parameters. **A,** Ramachandran plot of osteocalcin

structure. Red indicates most favored regions; yellow, additionally allowed

regions; beige, generously allowed regions; white, disallowed regions. It shows

that 93.3% of residues are in the most favored regions (A, B, L), 2 are in the

additionally allowed regions (a, b, l, p), none in the generously allowed (~a, ~b,

~l, ~p) and disallowed regions. **B,** indicates the position of each residues on the

Ramachandran map. This figure was generated with program PROCHECK in the

CCP4 package.

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	0	0.0	16.3	6.5	-2.5	BETTER
b. Chi-1 trans st dev	14	19.8	17.5	5.3	0.4	Inside
c. Chi-1 gauche plus st dev	11	10.7	16.0	4.9	-1.1	BETTER
d. Chi-1 pooled st dev	25	15.9	16.6	4.8	-0.1	Inside
e. Chi-2 trans st dev	7	10.8	19.3	5.0	-1.7	BETTER

Figure 5-3. Sidechain Parameters. Analysis of side chain conformations by the program PROCHECK.

Table 5-1. Refinement statistics of osteocalcin structure generated by the program CNS.

| Data Used in Refinement	
Resolution Range High (Å)	2.00
Resolution Range Low (Å)	27.72
Data Cutoff (σF)	3.0
Completeness (working+test) (%)	87.7
Number of Reflections	6230
Fit to data used in refinement	
R Value (Working set)	0.255
Free R Value	0.283
Free R value test set size (%)	15.2
B values	
Mean B Value	37.1
Estimated Coordination error	
ESD from Luzzati Plot (Å)	0.31
ESD from SigmaA (Å)	0.27
Rms deviation from ideal values	
Bond Lengths (Å)	0.006
Bond Angles (degrees)	1.3

Chapter 6

Crystal Structure of Porcine osteocalcin

General characterization of osteocalcin

Since the refined model consisted of residues 13 to 49, this structure will be referred to as POC_{13-49} hereafter. Porcine osteocalcin crystallized as a crystallographic dimer with a two fold symmetry about the b-axes (figure 6-1). There is no direct intermolecular protein-protein interaction within the dimer, but rather, the interactions that hold the dimer together are protein-Ca^{2+}-protein. The Gla residues on each monomer are arranged linearly on the protein surface and the row of Ca^{2+} is sandwiched between these two surfaces. The arrangement of the Ca^{2+} atoms is ordered and also has the same 2-fold relationship. The significance of this ordered arrangement will be discussed in chapter 8. The crystal structure POC_{13-49} consists of 3 helices, each is separated from another by a turn forming a helix-turn-helix-turn-helix motif. The first helix (H1) spans from Asp17 to Asn26 (figure 6-2A). All three Gla residues lie on one side of H1 helix with their side-chains radiating away from the protein core where they, together with Asp30, coordinate the 5 Ca^{2+} ions that form the dimer interface (figure 6-1).

The second helix (H2) spans from Asp28 to Asp34 (figure 6-2B). H2 is separated from H1 by a β-turn between Leu25 and Pro27, which is stabilized by the disulfide-bridge (figure 6-2D). The β-turn and the disulfide-bridge position H2 such that Asp30 is oriented correctly for participation in calcium chelation. The third helix (H3) spans from Phe38 to Tyr46 (figure 6-2C). H3 is turned back, via a turn between Ile36 and Phe38, to close proximity of H1 and H2; thereby, forming a hydrophobic core between the three helixes (figure 6-2D). The hydrophobic core is made up of residues Val22, Leu25, Leu32, Ala33, Ala41, Tyr42, Phe45 and Tyr46.
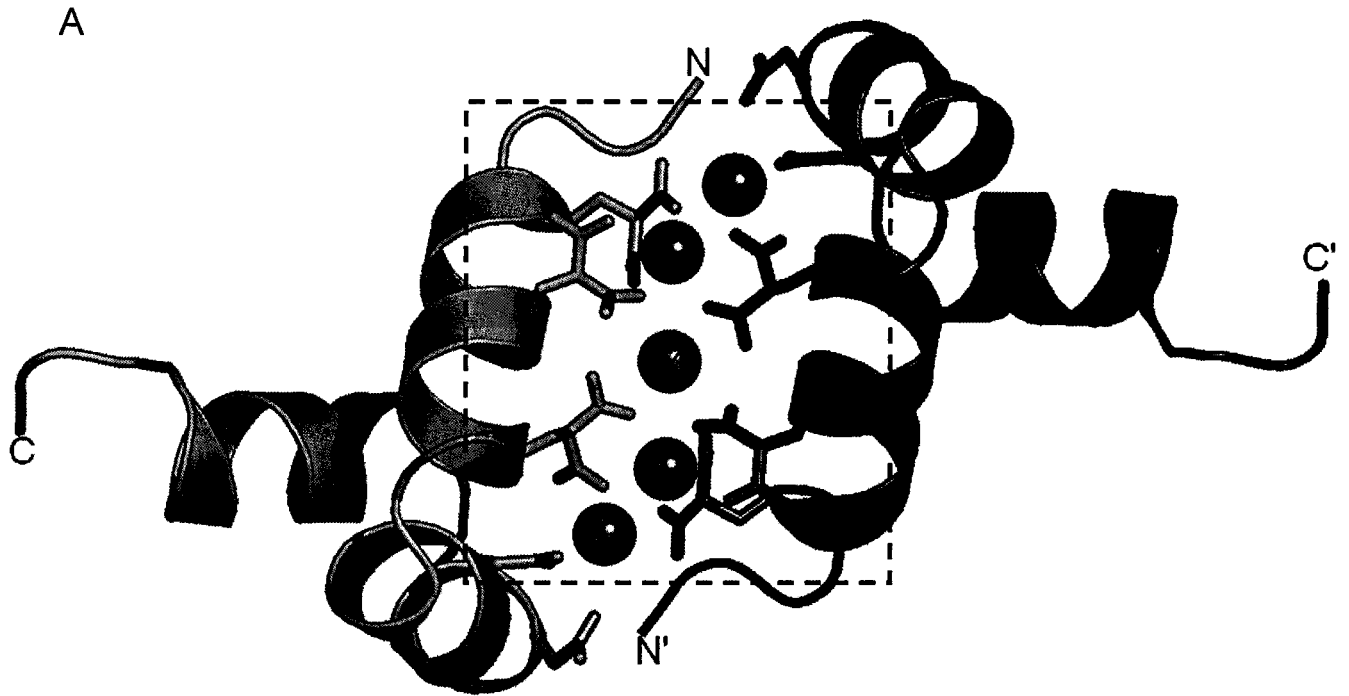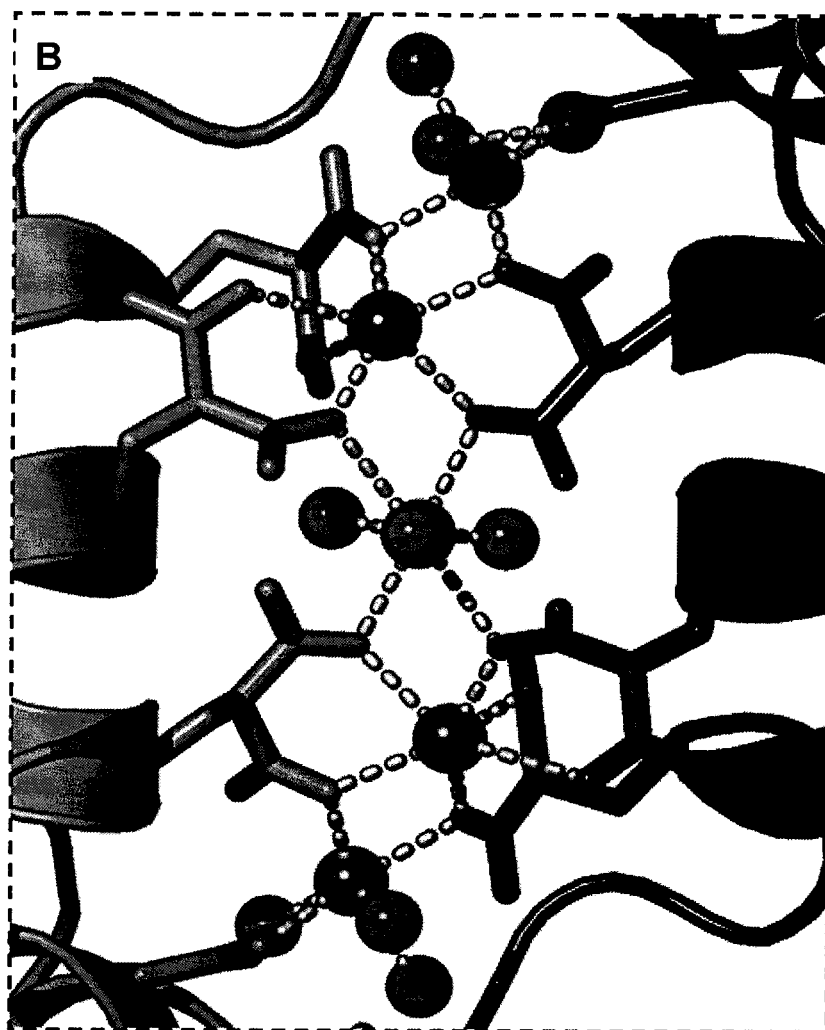
Figure 6-1. Crystal structure of osteocalcin dimer and the interaction at the dimer interface. **A**, Ribbon diagram of POC_{13-49} dimer as observed in the crystal. The black x on the center calcium indicates the position of a two-fold symmetry axis, which is perpendicular to the plane of the page. Side-chains of Gla residues and Asp30 are shown as stick model. **B**, Close-up view of the dimer interface. Purple spheres, calcium ions; red spheres, water molecules; yellow dotted lines, electrostatic interactions. The distances range from 2.08 Å to 2.98 Å, and average is 2.54 Å. Images were generated by Pymol.

A

H1

Leu26

Gla17

B

Ile36

H2

Pro27

C

H3

Phe38

Tyr46

D

Ile36

Gla17

E
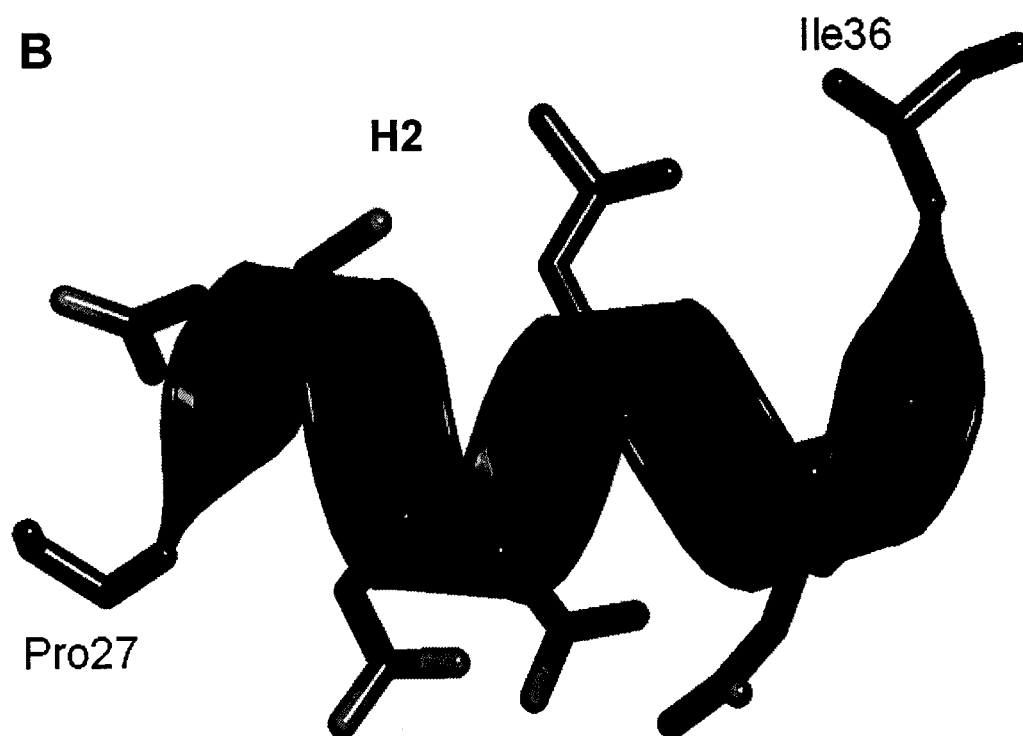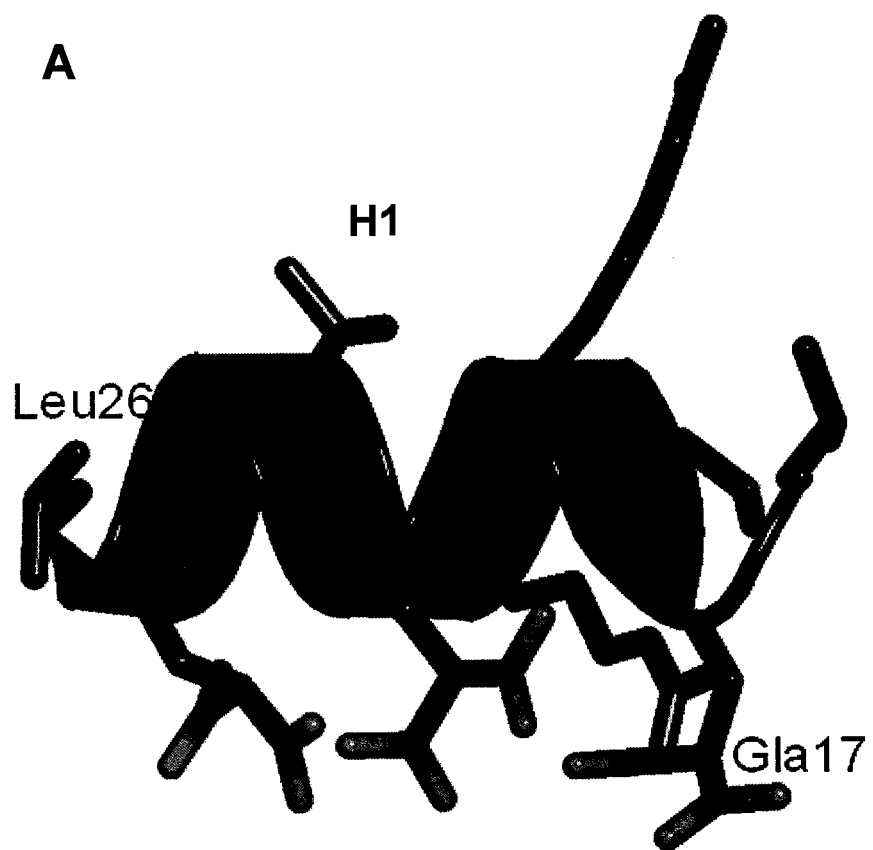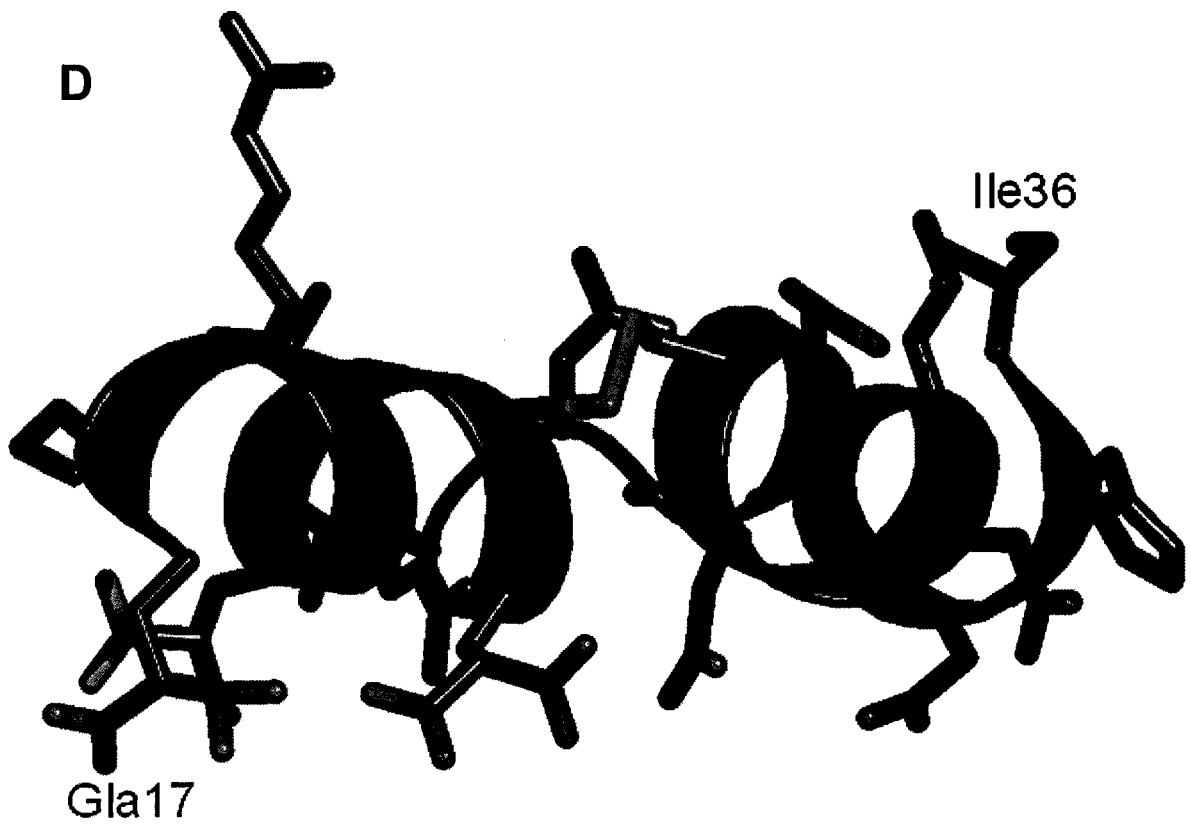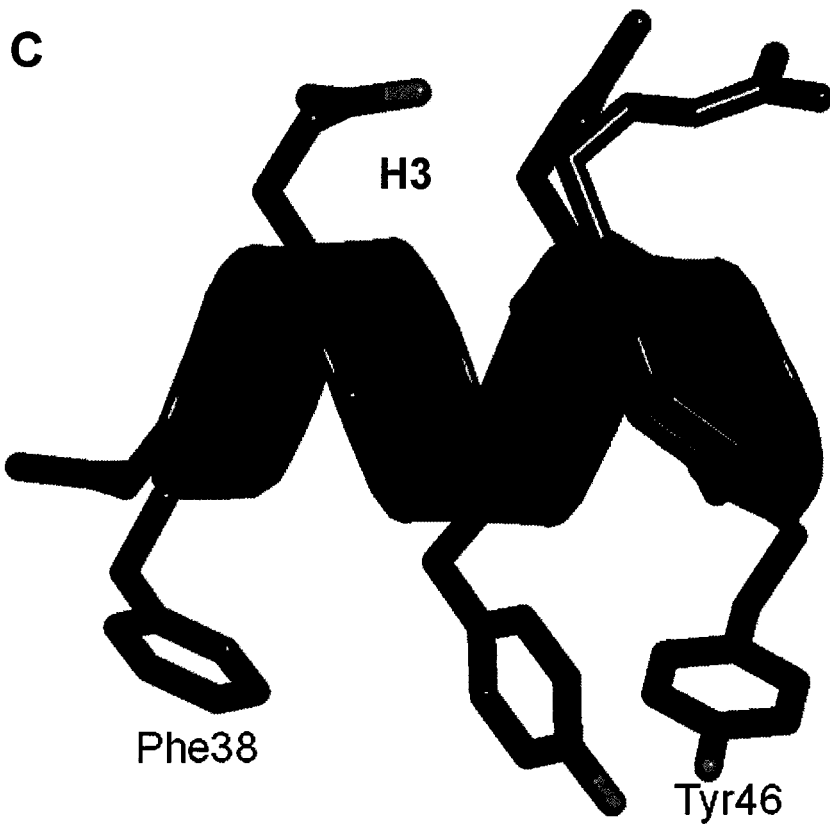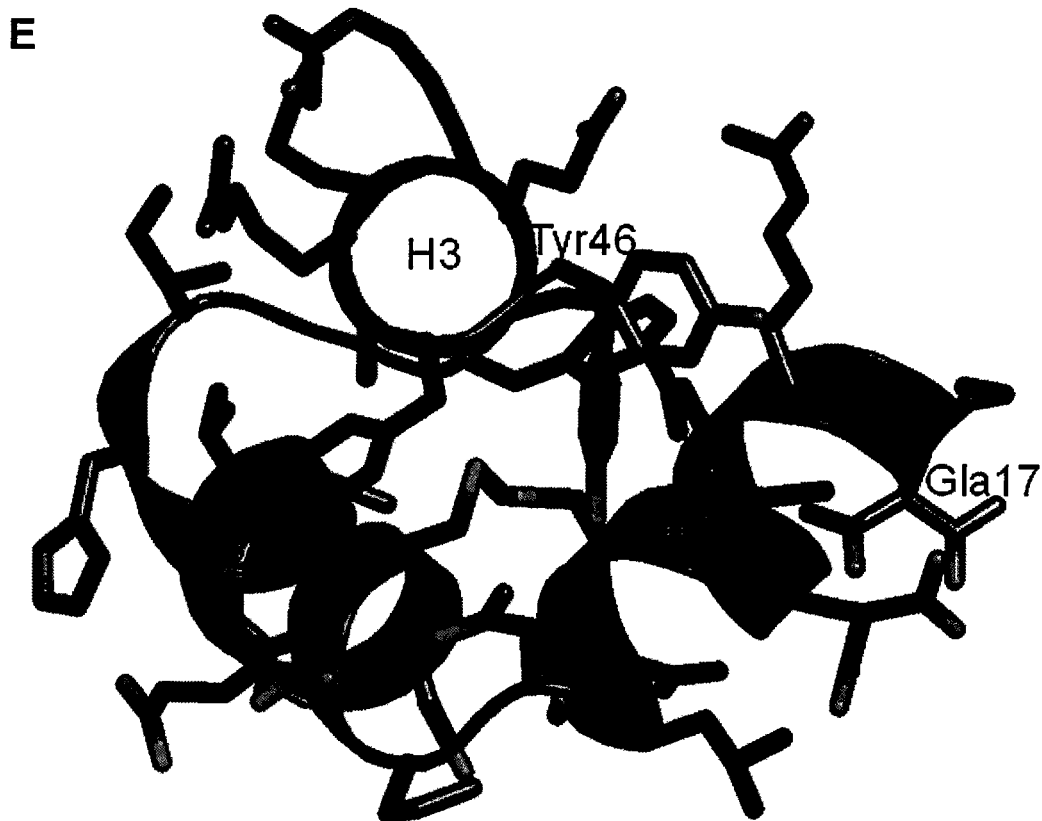
Figure 6-2. Secondary structures and their arrangement in osteocalcin. This

figure shows the ribbon diagrams of secondary structures found in osteocalcin.

Side chains are shown in stick model. **A**, H1; **B**, H2; **C**, H3; **D**, disulfide bridge

between H1 and H2; **E**, hydrophobic core formed between the three helices. The

backbone and all carbon atoms are colored green, where oxygen, nitrogen and

sulfur atoms are red, blue, and orange, respectively.
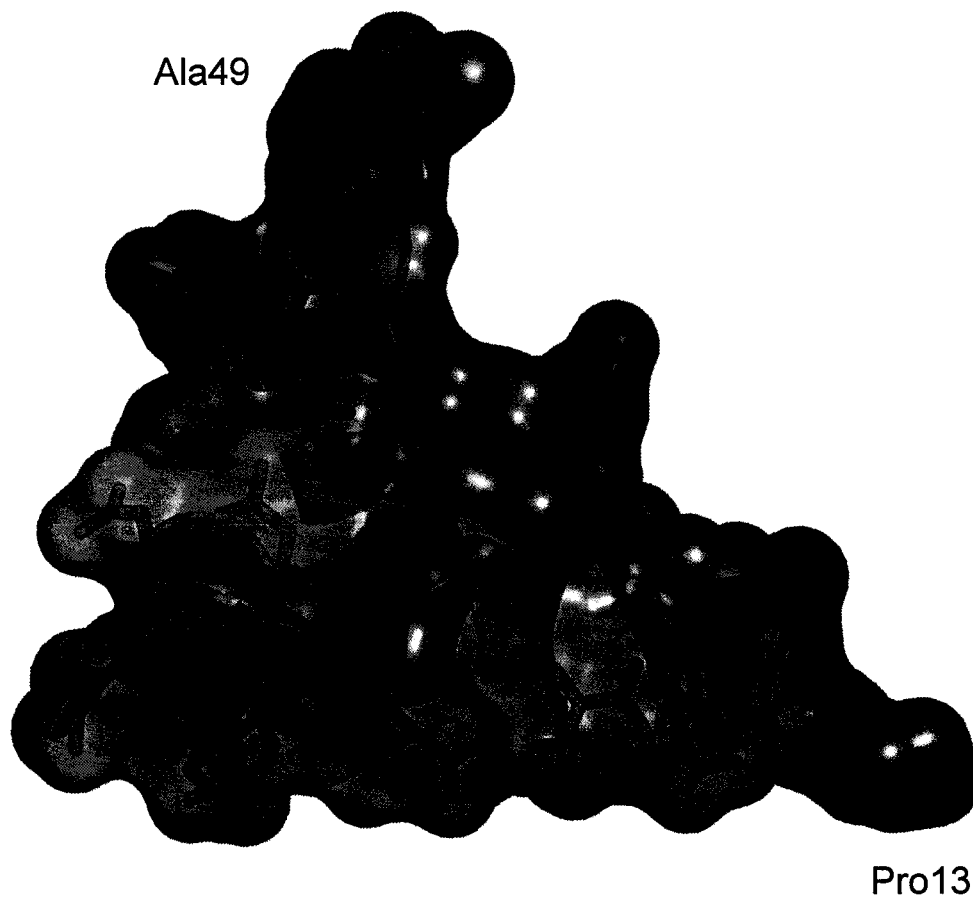
Figure 6-3. Distribution of charges on protein surface. Surface rendering of

POC$_{13\text{-}49}$ with the secondary structures revealed below the semitransparent

surface. Acidic residues are colored red, basic residues are colored green and

the remaining are colored gray. This figure was generated by PyMol.
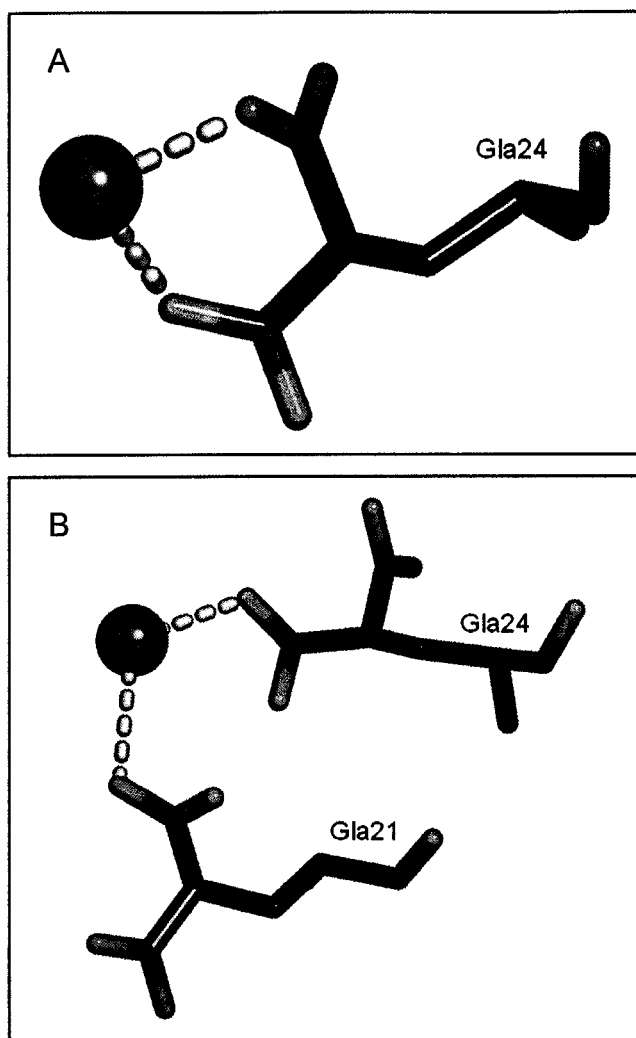
Figure 6-4. Calcium binding modes as observed in the crystal structure of osteocalcin. Side chains are shown in stick model, where carbon, oxygen, and nitrogen atoms are green, red, and blue, respectively. Yellow dotted lines represents electrostatic interactions. **A**, shows Gla24 interacting with calcium in malonate chelation mode. **B**, shows Gla21 and Gla24 interacts with calcium in a unidentate mode. Image was generated with PyMol.

```
                                                                    ___
 _____■■■■■■■■■___■■■■___■■■■■■■■_____

            1          10         20        30        40
Monkey     YLYQW LGAPA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Rabbit     QLING QGAPA PYPDP LEPKR EVCEL NPDCD ELADQ VGLQD AYQRF YGPV
Human      YLYQW LGAPA VYPDP LEPRR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Cow        YLDHW LGAPA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Pig        YLDHG LGAPA PYPDP LEPRR EVCEL NPDCD ELADH IGFQE AYRRF YGIA
Sheep      YLDPG LGAPA PYPDP LEPRR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Goat       YLDPG LGAPA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Dog        YLDSG LGAPV PYPDP LEPKR EVCEL NPNCD ELADH IGFQE AYQRF YGPV
Cat        YLAPG LGFPA PYPDP LEPKR EICEL NPDCD ELADH IGFQD AYRRF YGTV
Wallaby    YLYQT LGAPF PYPDP QENKR EVCEL NPDCD ELADH IGFSE AYRRF YGTA
Rat        YLNNG LGAPA PYPDP LEPHR EVCEL NPNCD ELADH IGFQD AYKRI YGTTV
Mouse      YL    GASV  PSPDP LEPTR EQCEL NPACD ELSDQ YGLKT AYKRI YGITI
Xenopus    SYGNN VGQGA AVGSP LESQR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Emu         SFAV GSSYG AAPDP LEAQR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Chicken HYAQDS GVAGA PYPDP LEPKR EVCEL NPDCD ELADH IGFQE AYRRF YGPV
Carp          AG TAPAD LTVAQ LESLK EVCEA NLACE HMMDV SGIIA AYTAY YGPIP Y
Tetraodon     AAGE PTLQQ LESLR EVCEL NIACD EMADP AGIVA AYAAY YGPPT F
Fugu          APGE PTPQQ LESLR EVCEL NIACD EMADT AGIVA AYAAY YGPPP F
Bluegill      AAGE LTLTQ LESLR EVCEA NLACE DMMDA QGIIA AYTAY YGPIP Y
Seabream      AAGQ LSLTQ LESLR EVCEL NLACE HMMDT EGIIA AYTAY YGPIP Y
Swordfish   A TRAGD LTPLQ LESLR EVCEL NVSCD EMADT AGIVA AYIAY YGPIQ F
```

Figure 6-5. Primary structures of osteocalcin. The sequences are arranged for best alignment. The corresponding secondary structures are indicated above the primary structures. H1, the first helix; H2, the second helix; H3 the third helix. Conserved residues are colored in red, functionally conserved residues are colored in blue. Monkey (Hauschka, Carr et al. 1982); Human (Poser and Price 1980); Cow (Price, Poser et al. 1976); Pig, goat, sheep, and wallaby (Huq, Tseng et al. 1984); Cat (Shimomura, Kanai et al. 1984); rat (Pan and Price 1985); mouse (Celeste, Rosen et al. 1986); rabbit (Virdi, Willis et al. 1991); chicken (Carr, Hauschka et al. 1981); swordfish (Price 1983); Carp, Fugu,

Discussion

As mentioned in chapter 1, the objective of this study was to investigate the interaction between osteocalcin and hydroxyapatite. Since osteocalcin crystallized as a dimer, the surfaces of the dimer were inspected for hydroxyapatite binding sites. In the dimer form, there was no surface that showed complimentarity to the flat and highly charged surface of hydroxyapatite. Furthermore, it has been shown that Gla residues interact directly with hydroxyapatite (Poser and Price 1979), in the dimer form the Gla residues are hidden in the core of the dimer and are not accessible for such interaction; therefore, osteocalcin is likely to function in its monomer form. In the monomeric form, the Gla residues are fully exposed on the surface of the protein and form a perfectly flat and highly charged surface, which can compliment well with calcium surfaces of hydroxyapatite. Detailed interaction between this surface and hydroxyapatite is discussed in chapter 8.

A striking feature of the structure is the ordered arrangement of calcium atoms in the dimer interface. Such order is characteristic of crystal lattices, suggesting that the calcium binding surface on osteocalcin may also be suited for binding to crystal surfaces.

The arrangement of negatively charged residues on the H1 surface is precise for calcium coordination in that the coordination geometries are near perfect. This may explain the fact that no other cations can substitute for

calcium's effect on osteocalcin folding (Atkinson, Evans et al. 1995) and

hydroxypaptite binding enhancement (Hauschka and Carr 1982). The average

ionic bonding distance between the protein and calcium ions is 2.54 Å, and the

coordination of each calcium ions showed octahedral geometry. Osteocalcin has

been shown to chelate calcium in the malonate chelation mode (involves a single

oxygen atom from each of the two carboxylate groups of the anion) (Mizuguchi,

Fujisawa et al. 2001). We found that POC_{13-49} chelates calcium in malonate mode

as well as in a unidentate manner (figure 6-4). It has been shown that calcium

binding increases helical content of OC (Hauschka and Carr 1982; Atkinson,

Evans et al. 1995) and induces dimer formation (Nousianinen, M., P. J. Derrick,

et al. 2002). From figure 6-1B, it can be seen clearly that calcium binding not only

neutralizes the negative charge repulsion between the Gla residues, the linking

of the Gla residues through calcium binding significantly stabilizes the H1 helix.

The folding of H1 helix provides a hydrophobic patch to complete the

hydrophobic core that holds the three helixes together; therefore, calcium binding

not only induces helix formation, it also induces overall protein folding. If

osteocalcin is functionally competent only when it is folded, then the fold-unfold

effect of calcium binding can serve as an on-off switch to regulate the function of

osteocalcin. Though it is tempting to imagine that such an on-off switch can be

used to restrict osteocalcin's function to bone, where local concentration of

calcium is high, but the concentration of calcium in the blood is in millimolar

range, at which osteocalcin should bind calcium. In any event, if osteocalcin's

biological function is to bind bone crystals and recruit bone processing cells, as currently believed, there must be a mechanism to restrict its activity to bone; whether such restriction mechanism involves calcium binding is not clear.

On the calcium-binding surface, residues Gla17, Gla21, Gla24, Asp30, and Asp34 are involved in calcium binding. If this surface is also involved in hydroxyapatite binding, these residues should be conserved in all species because they bind to the same crystal lattice. As shown in figure 6-5, these residues in osteocalcin from all species are fully conserved, with the exception of the three fishes (carp, blue gill, and seabream) where glutamic acid is substituted for aspartic acid at position 30, suggesting that the calcium binding surface is conserved in all species and further supports that it can be the hydroxyapatite binding surface. It has been shown that the C-terminal fragment (residues 45 to 49) alone can have chemotactic activity (Mundy and Poser 1983). The structure POC_{13-49} shows that the C-terminal projects away from the protein core on the opposite side of the calcium-binding surface and is free of steric hindrance for interaction with other factors. It is reasonable to imagine that osteocalcin binds to hydroxyapatite with its calcium-binding surface while exposing the C-terminal for interaction with bone processing cells, however, further experimental evidence is required to validate this view. Resolving such a picture is not only biologically important, it is also of biochemical interest in the sense that it is usual for such a small protein to consist of multi-functional domains. It is interesting that the amino acid sequence of the N-terminal "domain", which is not modeled in POC_{13-49}, is

relatively promiscuous with respect to the rest of the molecule. If the

hydroxyapatite-binding surface and the chemotactic "domain" are sufficient to

fulfill the observed duel function of osteocalcin, then it is unclear what role the N-

terminus plays with respect to the overall function of osteocalcin.

Chapter 7

Dynamic Light Scattering

Introduction

To determine whether osteocalcin forms dimer in solution as observed in our crystal structure, dynamic light scattering (DLS) was used. Dynamic light scattering is a quick and convenient method for characterizing hydrodynamic size and aggregation of the biomolecules (Cohen, Fisch et al. 1990). DLS is based on the change of interparticle constructive and destructive interference of scattered light as a consequence of the motion of the particles. In a simplified system that contains only two particles, when laser light strikes the two moving particles, each particle scatters part of the incident light in all directions. The interference of the two scattered waves is measured by a detector placed at 90 degrees from the incident beam. The distance between the two particles varies continuously in time; therefore, the interference pattern varies as well. When the distance between the two particles is equal to an integral multiple of the wavelength (distance = $n\lambda$), constructive interference occurs and maximum light intensity is detected. Conversely, when the interparticle distance is equal to the integral

multiple of half the wavelength (distance $= n\lambda/2$) destructive interference occurs

and results in minimum light intensity at the detector. The particles' motions are

inversely proportional to their sizes; therefore, the detected intensities fluctuate

more rapidly with smaller molecules, and conversely, more slowly with larger

molecules. The rate of this fluctuation is analyzed by a process known as

autocorrelation, where the intensity of scattered light is repetitively compared to

itself at increasing time intervals from an arbitrary zero time point. The

autocorrelation function ($R(\tau)$) is mathematically described as; $R(\tau) = \Sigma_t\, I_s(t)\, I_s(t$

$+ \tau)$ (Cohen, Fisch et al. 1990), where $I_s(t)$ is the scattered light intensity

measured at time t, $I_s(t + \tau)$is the scattered light intensity measured at a time

delayed from t by an increment τ in microsecond scale. From this equation, it can

be seen that when τ is small, $I_s(t)$ and $I_s(t + \tau)$are similar; therefore, the two

intensity values are autocorrelated and the value of $R(\tau)$ will be large. The

autocorrelation function is recalculated iteratively with incrementally larger values

of τ. As τ becomes increasingly larger, the values of $I_s(t)$ and $I_s(t + \tau)$becomes

increasingly different and eventually they are no longer autocorrelated. As a

result, the $R(\tau)$ function begins with its maximum value and decays to a minimum

value as τ increases. This decay is due to the random movement of the particles

in solution, known as Brownian movement. The speed of the Brownian

movement for particles with given shapes and sizes is dependent on the diffusion

coefficient, D. For a population of identical particles, the autocorrelation function

is exp($-Dq^2\tau$), where q is the magnitude of the scattering factor. Though DLS

measures D, the Stokes-Einstein relation is used to derive the hydrodynamic

radius (R_h) of the particle from D: $R_h = (k_BT)/(6\pi\eta D)$, where k_B represents

Boltzman's constant, T is the absolute temperature, and η is the viscosity of the

solvent. R_h is commonly referred as the size of the particle.

Methods

DLS was carried out using a DynaPro-801 molecular sizing instrument

equipped with a microsampler (Protein Solutions). Purified osteocalcin was made

3.5 mg/ml in 100 mM HEPES, pH 7.4. A 50 µl sample was passed through a

filtering assembly containing a 0.02 µm filter into a 20 µl chamber quartz cuvette.

The data were analyzed using the Dynamics 4.0 software. Measurements were

repeated for samples containing various concentration of calcium chloride (from

0 to 33.2 mM). Twenty measurements were recorded for each sample and

standard deviations from the mean were calculated with Microsoft Excel.

Results and discussion

As shown in figure 7-1, the hydrodynamic radius of osteocalcin as

determined by dynamic light scattering decreased with increasing concentration

of calcium chloride. Such a phenomenon can be the result of either protein folding or dimer dissociation. Since a larger molecule scatters more light than a smaller one, dissociation of a dimer would result in dramatic reduction of the overall scattered light intensity, whereas, conformational change has relatively little effect on the overall scattered light intensity. The overall scattered light intensity remained relatively constant over the range of calcium chloride concentrations tested, suggesting that the progressive reduction in the measured size was due to calcium dependent conformational change of osteocalcin. Such calcium dependent conformation change is consistent with the literature, where calcium was found to induce higher helical content in osteocalcin (Hauschka and Carr 1982; Isbell, Du et al. 1993; Atkinson, Evans et al. 1995). Osteocalcin appears to be mostly folded at 2 mM concentration of calcium, which is about the concentration of calcium in the extracellular fluid of mammals (Heaney 2002). At 2 mM $CaCl_2$, the molar ratio of calcium to osteocalcin is about 3.3. This implies that osteocalcin is mostly folded when bound to 3 calcium ions. This is similar to NMR studies which found that the calcium dependent folding of dog osteocalcin was virtually saturated at $[Ca^{2+}]$/[osteocalcin] ratio of 2.5. In any case, based on our dynamic light scattering results, there is no evidence of dimer formation under the conditions tested.
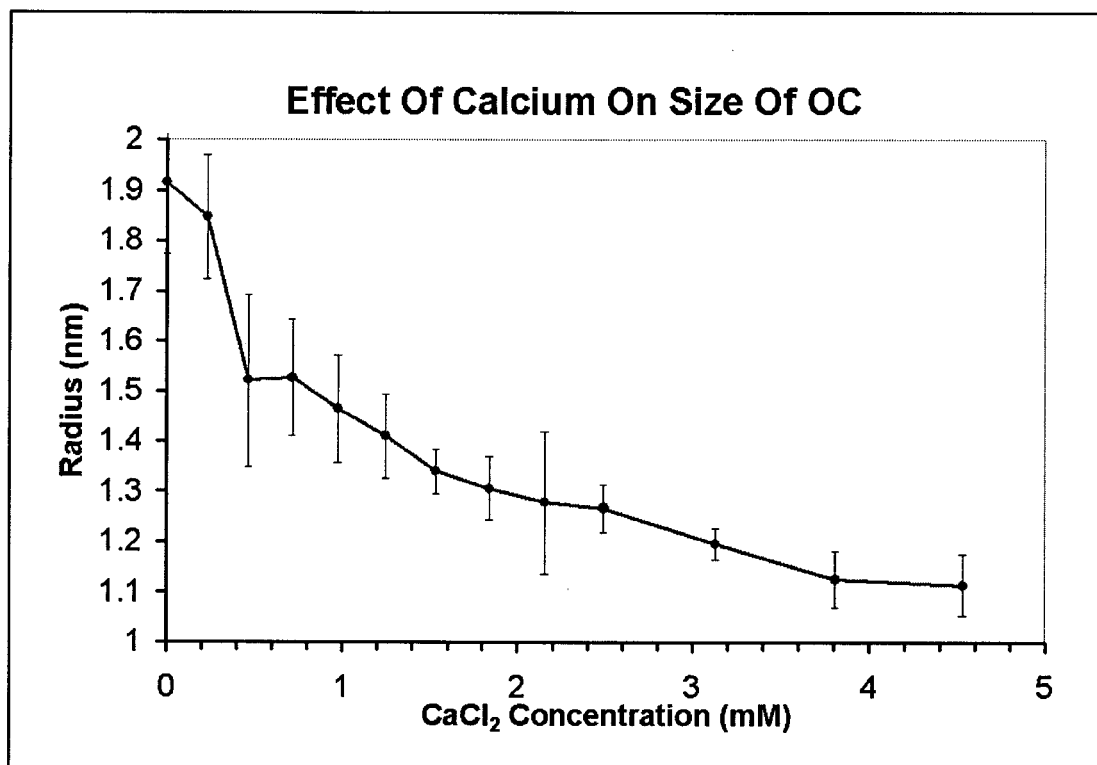
Figure 7-1. Effect of calcium on the radius of POC as measured by dynamic light scattering. POC (3.5 mg/ml in 100mM HEPES, pH7.4) in varying concentration of calcium chloride were assayed with DynaPro molecular sizing instrument (High Wycombe, UK). Twenty measurements were recorded for each sample and error bars are used to indicate the standard deviations from the mean.

Chapter 8

Osteocalcin binding to hydroxyapatite

Introduction

The structure of hydroxyapatite crystal has been determined and the coordinates are available at the online database. With the structure of osteocalcin determined, we now have the necessary components to investigate the interaction of osteocalcin with hydroxyapatite. Detailed analysis of protein-hydroxyapatite interaction can be easily accomplished by docking the atomic models of the two interacting surfaces. To do this, the interacting surfaces must first be identified. Identification of these surfaces is not trivial because there are virtually infinite possibilities. This chapter describes the procedure used to identify the putative binding surface on hydroxyapatite and the analysis of interactions between osteocalcin and hydroxyapatite.

Methods

Surface search

The coordinates of hydroxyapatite crystal were downloaded from the American Mineralogist Crystal Structure Database (www.geo.arizona.edu/AMS/amcsd.php) and its 3-dimentional atomic model was constructed with the program SURFACE (homemade). As mentioned in chapter 6, the arrangement of protein bound calcium ions showed precise geometrical order that resembled atomic placement in a crystal lattice. This order was undoubtedly dictated by the precise placement of the negatively charged residues on the Gla-containing surface of osteocalcin; therefore, if the Gla surface was to interact with hydroxyapatite, it would likely interact with a hydroxyapatite surface that consists of calcium atoms arranged geometrically similar to that of the interface calcium atoms. These assumptions reduced the enormous searched to a much smaller task. Instead of searching for complementary surfaces on the protein and hydroxyapatite, the task was reduced to the search of spatial match between a group of atoms. Furthermore, if the assumption that hydroxyapatite consists of calcium arranged similar to the protein bound calcium is correct, there are only two unique calcium positions in hydroxyapatite; therefore, each of the five calcium atoms in the search model (the protein bound calcium ions) must match to one of the two unique calcium

positions in hydroxyapatite. In this case, an atom from the search model can be manually placed into the position of one of the two unique calcium positions of hydroxyapatite (the pivot point) and the spatial match to the rest of the search model would involve only rotation of the search of the model, as a rigid body, about the pivot point. The search would be exhausted after the procedure is iterated for each of the five calcium ions in the search model. The "goodness of match" of each rotation was indicated by the root mean square distance (RMSD) between the search model and the calcium atoms in hydroxyapatite. The spatial search was carried out with the program ROTATE (homemade), where the rotational search was initially incremented by 5 degrees and RMSD for each rotation was calculated. The rotation results were sorted according to their corresponding RMSDs and top 50 rotations with the lowest RMSDs were selected for finer rotation increments. The process was iterated until no further improvement can be made with finer rotation increments. After the best match was identified, a thin slab of hydroxyapatite was constructed, with the matched atoms on the surface of the slab, for docking with osteocalcin.

Docking

Osteocalcin was docked to the surface of hydroxyapatite with the program LSQKAB (Kabsch 1976). Graphic display and analysis was done with the program PYMOL (DeLano 2002).

Results

To identify the putative hydroxyapatite crystallographic plane where POC

may selectively bind, we carried out a comprehensive search for a spatial match

between the POC-bound calcium to those on hydroxyapatite. The systematic

search yielded twenty-seven unique rotations with RMSD lower than 1.2. The

positions of these solutions are presented in figure 8-1. All twenty-seven

solutions matched to the same or symmetry related lattice points. These lattice

points are exposed on the prism face (100 and 010) of hydroxyapatite.

Consistent results were obtained in searches using five, four and three protein-

bound calcium ions as a search model. The atomic arrangement of osteocalcin-

bound calcium ions is nearly identical to those on the prism face of

hydroxyapatite found in our search (average RMSD=0.47); indicating that

osteocalcin may bind the prism face of hydroxyapatite. Superposition of a

calcium-bound osteocalcin monomer to the prism face of hydroxyapatite, with the

osteocalcin-bound calcium ions taking up the positions identified in our search,

showed excellent complimentarity of the two surfaces (figure 8-2). The

osteocalcin-calcium complex was docked to the surface of hydroxyapatite as a

rigid body and was lead solely by matching of the calcium atoms, no energy

minimization was performed. Surprisingly, osteocalcin docked to the surface of

hydroxyapatite beautifully with no steric hindrance at the protein-hydroxyapatite

interface. The superposition model showed the three central calcium ions (Ca2, Ca3, and Ca4) interact with the phosphate groups on the prism face of hydroxyapatite, while a γ-carboxyl group of Gla17 interacts with a geometrically selective calcium atom in the hydroxyapatite lattice (figure 8-2B).

C
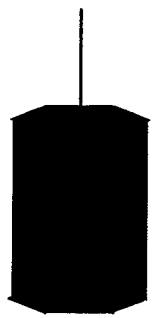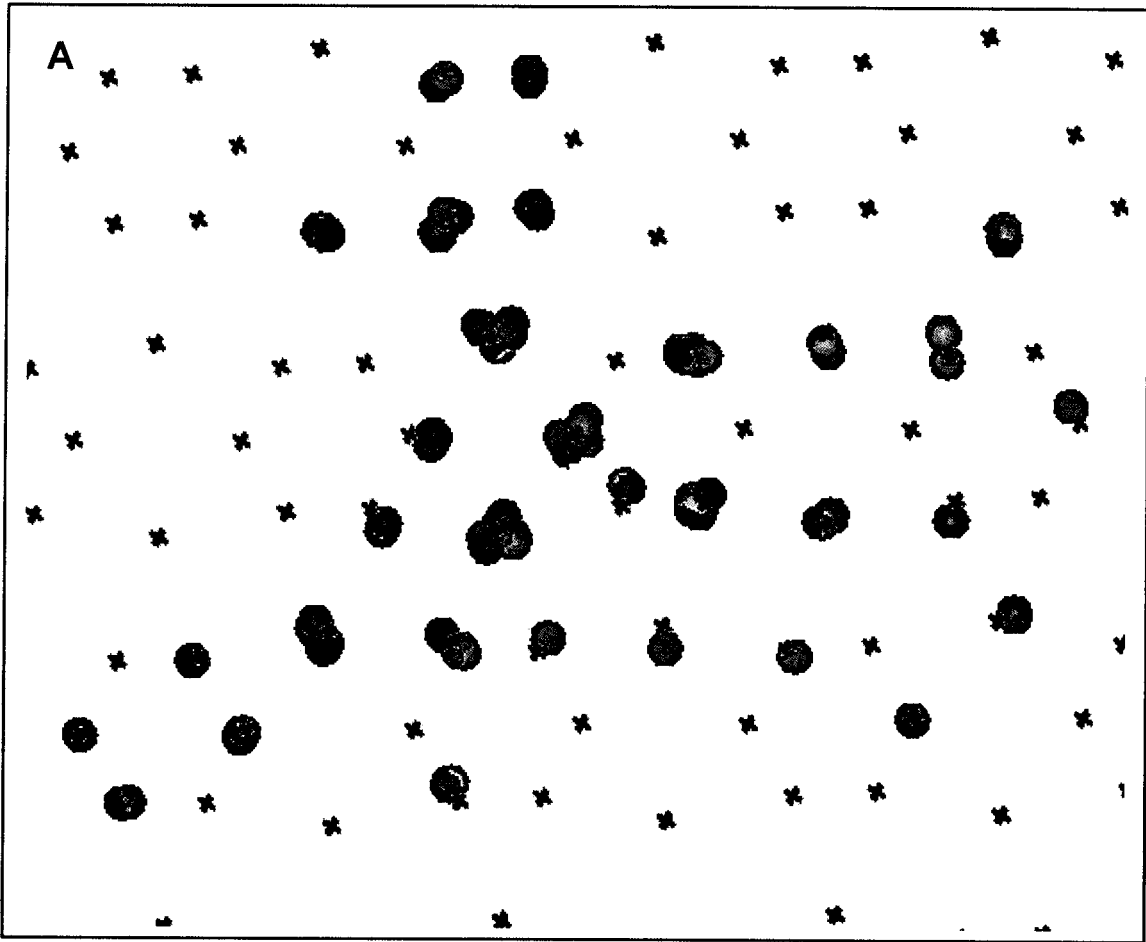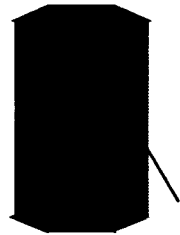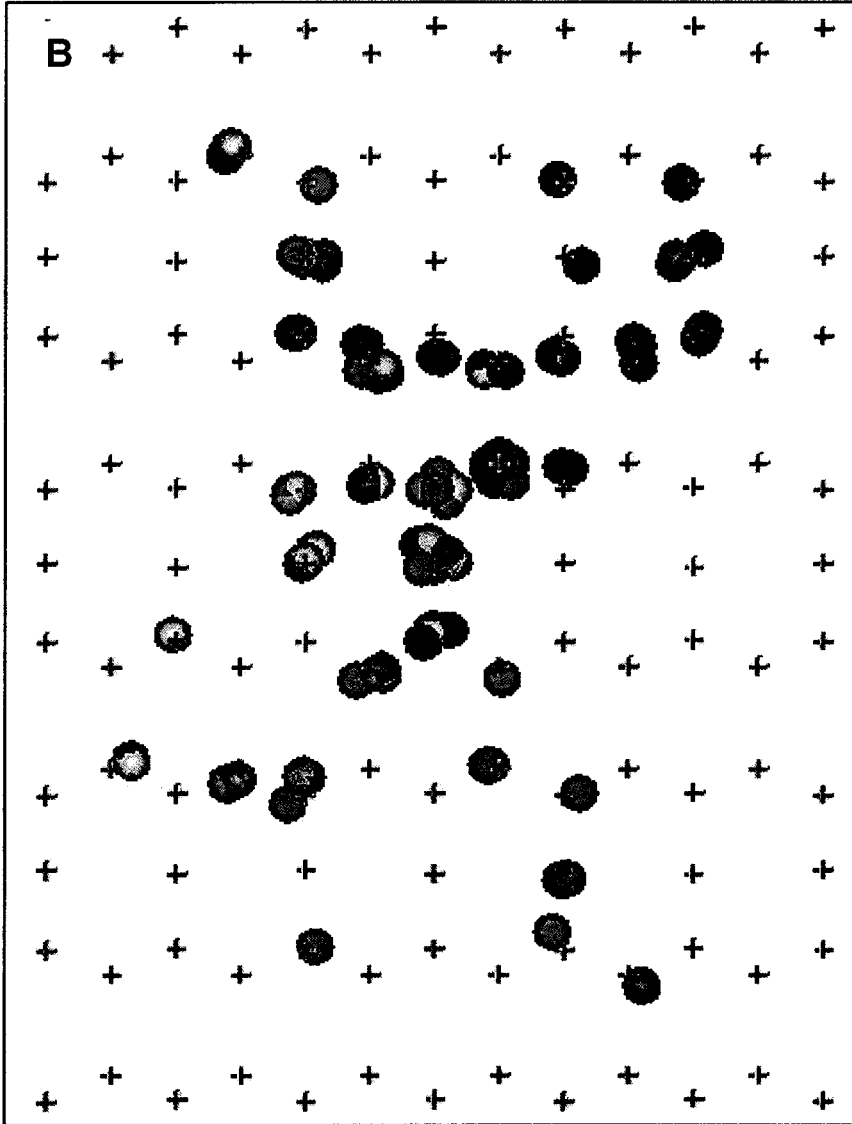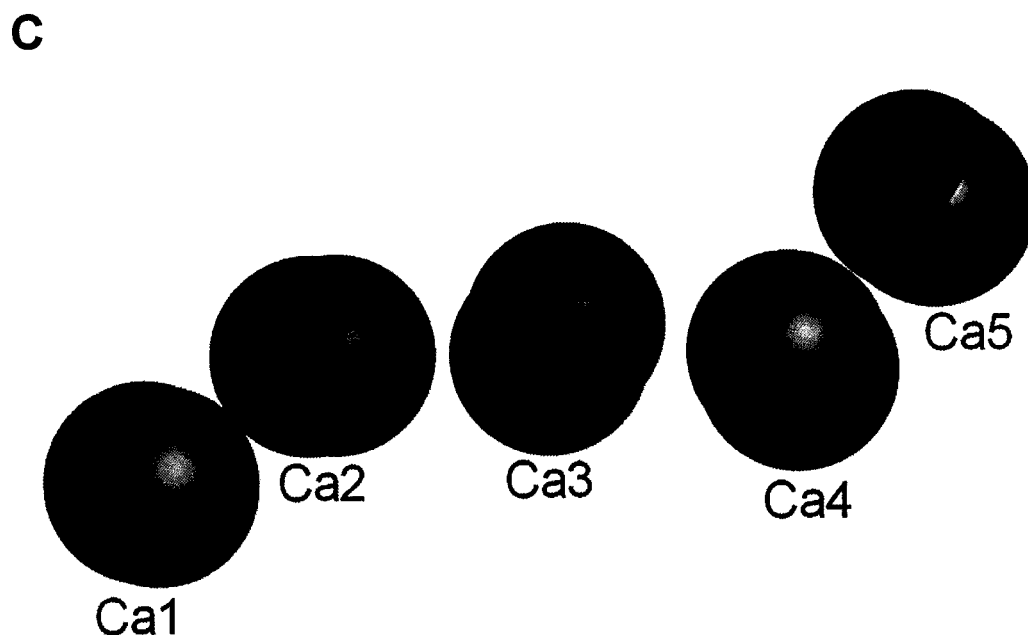
Figure 8-1. Atomic matches between hydroxyapatite and protein-bound calcium as found in the search. Colored spheres represent the 27 top solutions. Green crosses represent hydroxyapatite calcium atoms. **A**, viewed straight down the c-axis; **B**, viewed straight down the prism face; **C**, Space-fill presentation of POC_{13-49}-bound calcium alignment with hydroxyapatite calcium as found by our spatial search. Osteocalcin-bound calcium ions are arbitrarily numbered from left to right. Purple, POC_{13-49}-bound calcium; green, HA calcium. Distances between the center of purple and green are Ca1, 0.43 Å; Ca2, 0.45 Å; Ca3, 0.59 Å; Ca4, 0.51 Å; Ca5, 0.29 Å. Average r.m.s.d is 0.47 Å.

A
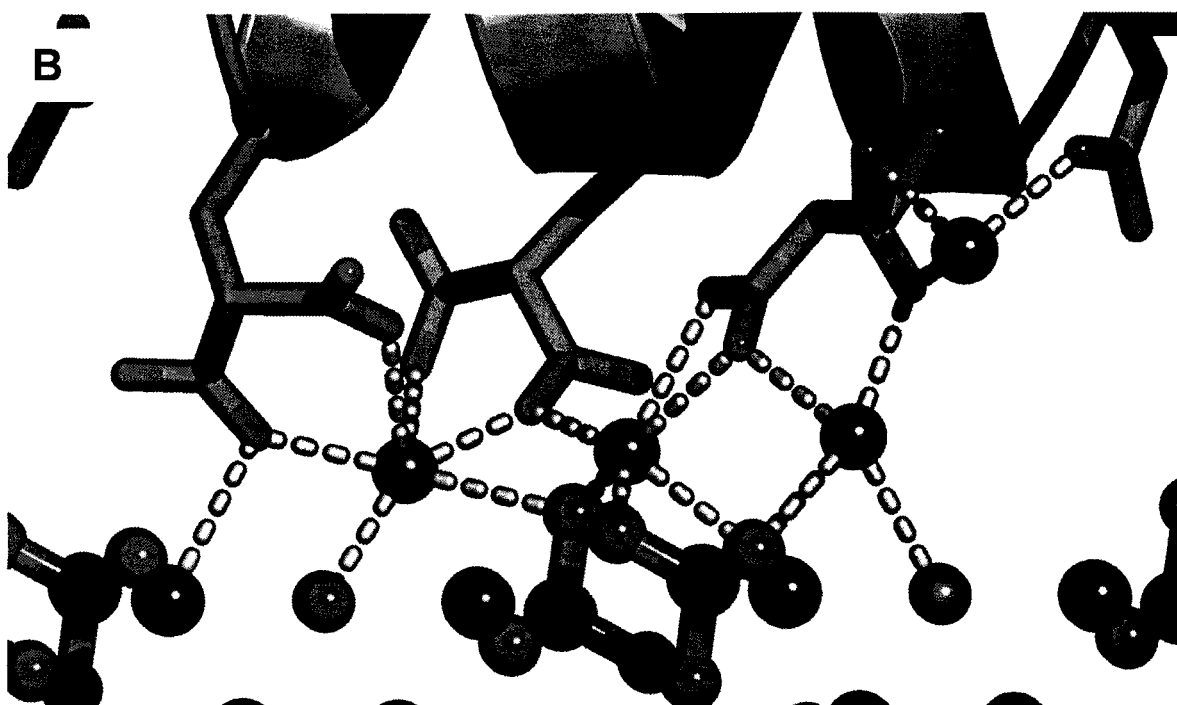
Figure 8-2. Osteocalcin binding onto the prism face hydroxyapatite. **A**, shows

POC$_{13\text{-}49}$ (orange ribbon) binding to calcium (purple spheres) which in turns bind

to the surface of HA (green). **B**, Shows the interactions between POC$_{13\text{-}49}$ and

HA. Distance of electrostatic interactions shown by yellow dotted lines range

from 1.97 Å to 2.89 Å; average, 2.45 Å. Orange, protein carbon; red, oxygen;

purple, protein-bound calcium; green, hydroxyapatite calcium; gray,

phosphorous; white, hydrogen. Distance calculation and image generation was

performed with PyMol.

Chapter 9

Discussion and Conclusion

In view of osteocalcin's abundance and its strong association with bone metabolism, deciphering its function would be instrumental in understanding bone biology, specifically bone metabolism. Indeed, elucidating osteocalcin's function has been the focus of many studies since its discovery over two decades ago. One of osteocalcin's functions is widely believed to bind bone crystals; thereby affecting the process of mineralization. This function was implicated by osteocalcin's abundance in mineralized bone and its high affinity to hydroxyapatite. One criticism of this putative function is that "one can not say that the function of osteocalcin is to bind hydroxyapatite simply because it is found to do so in vitro. The surface and coordination complimentarity must be considered". As shown in figure 8-2, the surface of the H1 helix showed excellent surface and coordination complimentarity with the prism face of hydroxyapatite. It should be noted that the docking of osteocalcin to the surface of hydroxyapatite was guided solely by the transformation that superimposed the protein-bound calcium ions, as a rigid-body, into the positions identified in the spatial search. No further movement or energy minimization was made after the transformation; therefore, the precise complementation was unanticipated and unbiased in that

regard. In bone, the hydroxyapatite crystal is a small and thin plate (50 x 25 x ~2 nm) with the c-axis expressed; therefore, the two longer dimensions (50 x 25 nm) define its prism face. This fact supports our model that osteocalcin binds to the prism face of hydroxyapatite, because it is the only face that is available for protein binding. Our model shows that osteocalcin binds to hydroxyapatite in its calcium-loaded form. Osteocalcin first binds free calcium ions in solution, which induces the protein to fold before it binds to hydroxyapatite. This theory is consistent with the observation that the presence of Ca^{2+} increased osteocalcin's affinity to hydroxyapatite (Poser and Price 1979). On the prism face of hydroxyapatite, the three protein-bound calcium ions interacts with the phosphate groups of hydroxyapatite while a carbonyl group of Gla17 interacts specifically with a calcium atom on the crystal surface. Such precise coordination highlights the essentiality of the Gla residues, since the mimicry of ordered apatite surface-phosphate on osteocalcin is not possible without γ-carboxylation. The superposition model also shows that when bound to the prism face of hydroxyapatite, Asp30 together with part of Gla24 binds to a calcium atom that is geometrically correct for the upper crystal layer. This suggests that osteocalcin may first bind to hydroxyapatite surface then induce a layer of crystal growth local to Asp30, or it may bind to the growth front as hydroxyapatite grows (crystals usually grow in layers). In any case, our docking experiment indicates that osteocalcin can bind specifically and precisely to the prism face of hydroxyapatite, which reinforces the speculated role of osteocalcin in binding to

bone crystals. Being bound to the crystal surface, it would undoubtedly inhibit crystal growth on the bound-surface. Such a mode of crystal growth inhibition has been extensively studied between antifreeze proteins and ice (Yang, Hon et al. 1998; Liou, Tocilj et al. 2000; Jia and Davies 2002). Since osteocalcin immobilizes a group of calcium ions and their geometrical arrangement is hydroxyapatite-like, such a protein-calcium complex can serve as a nucleus for hydroxyapatite growth; therefore, osteocalcin may also play a role in bone crystal nucleation. In any event, based on our model (figure 8-2) the Gla surface of osteocalcin is attached to the face of hydroxyapatite, while the C-terminal projects away from the core of the protein. This organization exposes the C-terminal where it can interact with bone processing cells or other factors. The role of the N-terminal is still puzzling. It is mostly hydrophobic and the sequence identity is variable among species. Since it is apparent that osteocalcin can fold into a stable structure independent of the N-terminal, it is probably not essential for overall protein folding. In view of the variability and hydrophobicity of the N-terminal, its function might be non-specific interaction with hydrophobic proteins in the extracellular matrix or membrane of bone cells.

The structural studies of osteocalcin presented in this thesis have implications for both the crystallographic techniques used to determine novel protein structures and for understanding the mechanisms of mineral crystal binding proteins.

The technique of single crystal, single-wavelength anomalous diffraction has rarely been practiced for the determination of protein structures. The results presented in this study clearly showed that very weak anomalous signal from relatively light atoms is sufficient to obtain an excellent experimental electrondensity map, into which the protein molecule can be unambiguously modeled. This attest to the general applicability of this technique for proteins with inherently bound metal atoms.

The dynamic light scattering results seem to have resolved the paradox of monomeric and dimeric form of osteocalcin in solution. While the crystal structure revealed that calcium binding mode of osteocalcin involves unidentate chelation mode as well as in malonate mode that was previously found (Mizuguchi et al. 2001).

A major paradox in the field of biomineral research is whether osteocalcin's biological function involves mineral crystal binding, and the effect of such binding. In contrast to the previously predicted binding model where osteocalcin binds to the basal plane of hydroxyapatite, the crystal structure of osteocalcin described here clearly showed that its Gla-containing surface could bind to the prism face of hydroxyapatite with high specificity, it seems reasonable that the presence of bound osteocalcin will inhibit further growth on this surface. Furthermore, since osteocalcin interacts with hydroxyapatite mainly via the bound calcium atoms, it seems fair to infer that the binding mechanism involves calcium binding to osteocalcin and induces protein folding before attaching to the

crystal surface. The data presented here can be used as a working model for further investigations as well as development of bone targeting drugs. Other bone mineral binding agents, specifically bisphosphonates, are established effective drugs in the treatment of osteoporosis (Watts, Harris et al. 1990; Liberman, Weiss et al. 1995), breast cancer metastasis (Hortobagyi, Theriault et al. 1996), and other bone diseases (Orcel and Beaudreuil 2002). Osteocalcin is a natural bone-binding agent and is undoubtedly important in bone metabolism, understanding its precise biological role would likely have therapeutic implications.

References

Atkinson, R. A., J. S. Evans, et al. (1995). "Conformational studies of osteocalcin in solution." Eur J Biochem **232**: 515-521.

Bijvoet, J. M. (1949). Proc. Acad. Sci. Amsterdam **B52**: 313.

Bijvoet, J. M. (1954). Nature **173**: 888.

Bodine, P. V. and B. S. Komm (1999). "Evidence that conditionally immortalized human osteoblasts express an osteocalcin receptor." Bone **25**(5): 535-543.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: a new software suite for macromolecular structure determination." Acta Crystallogr **D 54**: 905-921.

Cancela, M. L., M. K. Williamson, et al. (1995). "Amino-acid sequence of bone Gla protein from the African clawed toad Xenopus laevis and the fish Sparus aurata." Int. J. Pept. Protein Res. **46**(5): 419-423.

Carr, S. A., P. V. Hauschka, et al. (1981). "Gas chromatographic mass spectrometric sequence determination of osteocalcin, a gamma-carboxyglatamic acid containing protein of chicken bone." J. Biol. Chem. **256**: 9944-9950.

Celeste, A. J., V. Rosen, et al. (1986). "Isolation of the human gene for bone Gla protein utilizing mouse and rat cDNA clones." EMBO J. **5**: 1885-1890.

Chenu, C., S. Colucci, et al. (1994). "Osteocalcin induces chemotaxis, secretion of matrix proteins, and calcium-mediated intracellular signaling in human osteoclast-like cells." J Cell Biol **127**(4): 1149-1158.

Cohen, D. E., M. R. Fisch, et al. (1990). "Principles of Laser Light-scattering Spectroscopy: Applications to the physicochemical Study of Model and Native Biles." Hepatology **12**: 113S-122S.

Colombo, G., P. Fanti, et al. (1993). "Isolation and complete amino acid sequence of osteocalcin from canine bone." J Bone Miner Res **8**(6): 733-743.

Compton, J. E. (2002). "Bone marrow and bone: a functional unit." J. Endocrinology **173**: 387-394.

Conn, K. M. and J. D. Termine (1985). "Matrix protein profiles in calf bone development." Bone **6**(33-36): 1985.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, DeLano Scientific.

Driessens, F. C. M. and R. M. H. Verbeeck, Eds. (1990). Biominerals. Boca Raton, CRC Press.

Ducy, P., C. Desbois, et al. (1996). "Increased bone formation in osteocalcin-deficient mice." Nature **382**(6590): 448-452.

Gallop, P. M., J. B. Lian, et al. (1980). "Carboxylated calcium-binding proteins and vitamin K." N. Engl. J. Med. **302**: 1460-1466.

Hauschka, P. V. and S. A. Carr (1982). "Calcium-dependent a-heical structure in osteocalcin." Biochemistry **21**: 2538-2547.

Hauschka, P. V., S. A. Carr, et al. (1982). "Primary structure of monkey osteocalcin." Biochemistry **21**: 638-642.

Hauschka, P. V., J. Frenkel, et al. (1983). "Presence of osteocalcin and related higher molecular weight 4-carboxyglutamic acid-containing proteins in developing bone." J. Biol. Chem. **258**: 176-182.

Hauschka, P. V., J. B. Lian, et al. (1989). "Osteocalcin and Matrix Gla Protein: Vitamin K-Dependent Proteins in Bone." Physiol Rev **69**(3): 990-1047.

Hauschka, P. V., J. B. Lian, et al. (1975). "Direct identification of the calcium-binding amino acid gamma-carboxyglutamate in mineralized tissue." Natl. Acad. Sci. USA **72**: 3925-3929.

Hauschka, P. V. and M. L. Reid (1978). "Timed Appearance of a Calcium-Binding Protein Containing g-Carboxyglutamic Acid in Developing Chick Bone." Dev Biol **65**(2): 431-436.

Hauschka, P. V. and M. L. Reid (1978). "Vitamin K dependence of a calcium binding protein containing gamma-carboxyglutamic acid in chicken bone." J. Biol. Chem. **253**: 237-243.

Heaney, R. P. (2002). Calcium. Principles of Bone Biology. J. P. Bilezikian, L. G. Raisz and G. A. Rodan. San Diego, Academic Press. **2**: 1325-1336.

Hendrickson, W. A., J. L. Smith, et al. (1985). "Direct Phase Determination Based on Anomalous Scattering." Methods Enzymol **115**: 41-55.

Hortobagyi, G. N., R. L. Theriault, et al. (1996). "Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases." N Engl J Med **335**(24): 1785-1791.

Huq, N. L., A. Tseng, et al. (1984). "The amino acid sequences of goat, pig and wallaby osteocalcins." Biochem. Int. **8**: 521-527.

Isbell, D. T., S. Du, et al. (1993). "Metal Ion Binding to Dog Osteocalcin Studied by ^{1}H NMR Spectroscopy." Biochemistry **32**: 11352-11362.

Jia, Z. and P. L. Davies (2002). "Antifreeze proteins: an unusual receptor-ligand interaction." Trends Biochem Sci **27**(2): 101-106.

Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr **A 47**: 110-119.

Kabsch, W. (1976). "A solution for the best rotation to relate two sets of vectors." Acta Crystallogr **A 32**: 922-923.

Lian, J. B., A. H. Roufosse, et al. (1982). "Concentration of osteocalcin and phosphoprotein as a function of mineral content and age in cortical bone." Calcif. Tissue Int. **34**(Suppl. 2): 82-87.

Liberman, U. A., S. R. Weiss, et al. (1995). "Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis." N Engl J Med **333**: 1437-1443.

Liou, Y. C., A. Tocilj, et al. (2000). "Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein." Nature **406**(6793): 322-324.

McPherson, A. (1989). Preparation and analysis of protein crystals. Malabar, Krieger Publishing Co.

Mizuguchi, M., R. Fujisawa, et al. (2001). "Fourier-Transform Infrared Spectroscopic Study of Ca^{2+}-Binding to Osteocalcin." Calcif Tissue Int **69**: 337-342.

Mundy, G. R. and J. W. Poser (1983). "Chemotactic activity of the gamma-carboxyglutamic acid-containing protein in bone." Calcif. Tissue Int. **35**: 164-168.

Nishimoto, S. K., N. Arake, et al. (1992). "Discovery of bone gamma-carboxyglatamic acid protein in mineralized scales. The abundance and structure of Lepomis macrochirus bone gamma-carboxyglutamic acid protein." J. Biol. Chem. **267**(16): 11600-11605.

Nishimoto, S. K., J. H. Waite, et al. (2003). "Structure, Activity, and Distribution of Fish Osteocalcin." J. Biol. Chem. **278**(14): 11843-11848

Orcel, P. and J. Beaudreuil (2002). "Bisphosphonates in bone diseases other than osteoporosis." Joint Bone Spine **69**: 19-27.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol **276**: 307-326.

Pan, L. C. and P. A. Price (1985). "The properties of rat bone gamma-carboxyglutamic acid protein shares homology with other vitamin K dependent protein precursors." Proc. Natl. Acad. Sci. USA **82**: 6109-6113.

Pastoureau, P., P. Vergnaud, et al. (1993). "Osteopenia and bone-remodeling abnormalities in warfarin-treated lambs." J Bone Miner Res **8**(12): 1417-1426.

Poser, J. W. and P. A. Price (1979). "A Method for Decarboxylation of g-Carboxyglutamic Acid in Proteins." J Biol Chem **254**(2): 431-436.

Poser, J. W. and P. A. Price (1980). "Isolation and sequence of the vitamin K-dependent protein human bone: undercarboxylation of the first glutamic acid residue." J. Biol. Chem. **255**: 8685-8691.

Price, P. A. (1983). Osteocalcin. Bone and Mineral Research. W. A. Peck. Amsterdam, Excerpta Med. **1**: 157-190.

Price, P. A., J. W. Poser, et al. (1976). "Primary structure of the gamma-carboxyglutamic acid-containing protein from bovine bone." Proc. Natl. Acad. Sci. USA **73**: 3374-3375.

Price, P. A. and M. K. Willamson (1981). "Effects of warfarin on bone: studies on the vitamin K-dependent protein of rat bone." J. Biol. Cheml. **256**: 12754-12759.

Price, P. A., M. K. Williamson, et al. (1982). "Excessive mineralization with growth plate closure in rats on chronic warfarin treatment." Proc. Natl. Acad. Sci. USA **79**: 7734-7738.

Read, R. J. (1986). "Improved Fourier coefficients for maps using phases from partial structures with errors." Acta. Cryst. **A42**: 140-149.

Shimomura, H., H. Kanai, et al. (1984). "Primary structure of cat osteocalcin." J. Biochem. Tokyo **96**: 405-411.

Terwilliger, T. C. (1999). "Maximum-likelihood density modification." Acta Crystallogr **D 56**: 1863-1871.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallogr **D 55**: 849-861.

Virdi, A. S., A. C. Willis, et al. (1991). "Primary aminoacid sequence of rabbit osteocalcin." Biochem Soc Trans **19**(4): 373S.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol. H. W. Wychoff, C. H. W. Hirs and S. N. Timasheff. Orlando, Academic Press. **115:** 90-112.

Nousianinen, M., P. J. Derrick, et al. (2002). "A Mass Spectrometric Study of Metal Binding to Osteocalcin." Chemistry & Biology **9**: 195-202

Watts, N. B., S. T. Harris, et al. (1990). "Intermittent cyclical etidronate treatment of postmenopausal osteoporosis." N Engl J Med **323**: 73-79.

Weiner, S., W. Traub, et al. (1999). "Lamellar Bone: Structure-Function Relations." J Struct Biol **126**: 241-255.

Weiner, S. and H. D. Wagner (1998). "The material bone: structure-mechanical function relations." Annu. Rev. Mater. Sci. **28**: 271-298.

Yang, D. S., W. C. Hon, et al. (1998). "Identification of the ice-binding surface on a type III antifreeze protein with a "flatness function" algorithm." Biophys J **74**(5): 2142-2151.

Ziv, V., H. D. Wagner, et al. (1996). "Microstructure-Microhardness relations in parallel-fibered and lamellar bone." Bone **18**(5): 417-428.

Appendix 1

Screen conditions that yielded crystals of porcine osteocalcin

Results of crystal screening with porcine osteocalcin, 12 mg/ml in 100mM

HEPES pH 7.5. Each hanging drop contains 2µl of protein solution, 1µl of 25mM

CaCl2 in 100mM HEPES, and 2µl of Hampton Research screen formulation.

Each well contains 500ml of Hampton Research screen formulation.

Hampton Research
Screen I

Condition	*Observation*
#4 **Salt**: None **Buffer**: 0.1M Tris-HCl pH 8.5 **Precipitant**: 2.0M $(NH4)_2SO_4$	- Isolated rods - Some clusters - Phase separation
#15 **Salt**: 0.2M $(NH4)_2SO_4$ **Buffer**: 0.1M Sodium Cacodylate pH 6.5 Precipitant: 30% w/v PEG 400	- Clusters of needles

#17
Salt: 0.2M Lithium Sulfate
        monohydrate
Buffer: 0.1M Tris-HCl pH 8.5
Precipitant: 30% PEG 4000
	- Some isolated needles
	- Some clusters of
	  needles

#22
Salt: 0.2M Sodium Acetate
        trihydrate
Buffer: 0.1M Tris-HCl pH 8.5
Precipitant: 30% PEG 4000
	- Some isolated needles
	- Some clusters of
	  needles

#25 **Salt**: None **Buffer**: 0.1M Imidazole pH 6.5 **Precipitant**: 1.0M Sodium	- Isolated rods

Acetate trihydrate	
#36 **Salt**: None **Buffer**: 0.1M Tris-HCl pH 8.5 **Precipitant**: 8% w/v PEG 800	- Some isolated needles - Some clusters of needles

#39
Salt: None
Buffer: 0.1 M HEPES-Na
    pH 7.5
Precipitant: 2% v/v PEG 400
        2.0M Ammonium
        Sulfate
	- Some isolated needles
	- Some clusters of needles

#40
Salt: None
Buffer: 0.1M Tri-Sodium
    Citrate dihydrate pH
5.6
Precipitant: 20% v/v
isopropanol
        20% w/v PEG
4000
	- Clusters of micro crystals

#41 **Salt**: None **Buffer**: 0.1M HEPES-Na pH 7.5 **Precipitant**: 10% v/v isopropanol 20% w/v PEG 4000	- Big isolated needles with some clusters
#42 **Salt**: 0.05M mono-Potassium dihydrogen phosphate **Buffer**: None **Precipitant**: 20% w/v PEG 8000	- Many small isolated rods

Hampton Research
Screen II

Condition	Observation
#1 **Salt**: 2.0M NaCl **Buffer**: None **Precipitant**: 10% w/v PEG 6000	- Showers of micro crystals and some clusters of needles
#8 **Salt**: 1.5M NaCl **Buffer**: None **Precipitant**: 10% v/v Ethanol	- Needles

#13
Salt: 0.2M Ammonium Sulfate
Buffer: 0.1M Sodium Acetate
        pH 4.6
Precipitant: 30% PEG mono-
            methylether 2000
	- Not very defined shape, rock-like

#21
Salt: 0.1M Na Phosphate mono-
     basic
     0.1M K Phosphate mono-
     basic
Buffer: 0.1M MES pH 6.5
Precipitant: 2.0M NaCl
	- Not very defined shape, rock-like

#22 **Salt**: None **Buffer**: 0.1M MES pH 6.5 **Precipitant**: 12% w/v PEG 20,000	- Clusters of messy looking crystals
#26 **Salt**: 0.2M Ammonium Sulfate **Buffer**: 0.1M MES pH 6.5 **Precipitant**: 30% w/v PEG monomethylether 5000	- Clusters of needles
#29 **Salt**: 0.5M Ammonium Sulfate **Buffer**: 0.1M HEPES pH 7.5 **Precipitant**: 30% v/v MPD	- Not very defined shape, rock-like

#30 **Salt**: None **Buffer**: 0.1M HEPES pH 7.5 **Precipitant**: 10% PEG 6000 5% w/v MPD	- Large needles
#32 **Salt**: 0.1M NaCl **Buffer**: 0.1M HEPES pH 7.5 **Precipitant**: 1.6M Ammonium Sulfate	- Isolated and clusters of needles. - Some phase separation - Observed after one week
#37 **Salt**: None **Buffer**: 0.1M HEPES pH 7.5 **Precipitant**: 10% w/v PEG 8000 8% v/v Ethylene glycol	- Large rods - Seems to be growing too fast
#38 **Salt**: None **Buffer**: 0.1M HEPES pH 7.5 **Precipitant**: 20% PEG 10,000	- Large rods - Seems to be growing too fast
#40 **Salt**: 0.1M CaCl$_2$ ·2H$_2$O **Buffer**: 0.1M HEPES pH 7.5 **Precipitant**: 1.6M Ammonium Sulfate	- Nice large rods, but a little dried up
#42 **Salt**: 1.5M Ammonium Sulfate **Buffer**: 0.1M Tris pH 8.5 **Precipitant**: 12% v/v glycerol	- Many tiny needles
#43 **Salt**: 0.2M Ammonium Sulfate monobasic **Buffer**: 0.1M Tris pH 8.5 **Precipitant**: 50% w/v MPD	- Long small needles
#44 **Salt**: None **Buffer**: 0.1M Tris pH 8.5 **Precipitant**: 20% v/v Ethanol	- One large triangular rod, grew after one week - One large hexagon rod, grew after two weeks

#48 **Salt**: 2% v/v Dioxane **Buffer**: 0.1M Bicine pH 9.0 **Precipitant**: 10% w/v PEG 20,000	- Long small needles

Appendix 2

Macros for data processing with the HKL package (Denzo and

Scalepack)

Indexing

[modified by Quyen Hoang for processing osteo. native, collected at Andy Howard's beam line]
[on mar 165m ccd detector Aug 2000]
[This script includes site.dat, experiment.dat, and auto.dat]

[Site Template auto-indexing file for operation of DENZO McMaster University]

[site.dat]

format ccd unsupported-m165   [detector type]

wavelength   1.70000
plarization -0.03   [I've copied this from Andy Howard's denzo.in file]
cassette rotx -0.207 roty   0.307 rotz   0.000   [orientation of the film cassette]
spot radius 0.50
background radius 0.60
ibox 16 16
overlap spot   [Flags overlapped spots]
x beam 81.6   [I've copied this from Andy Howard's denzo.in file]
y beam 83.3
crossfire y -0.004   [Measure of the X-ray beam divergence and focusing,]
crossfire x 0.018   [if beam is parallel or focused at on the detector]
crossfire xy -0.068   [then these values are 0]
profile fitting radius 20.0   [Default value is 20.0]
weak level 5.0   [Intensity rejection level, I/sigma ratio]
[Experiment.dat]
title 'osteocalcin aps osn31'
oscillation range 1.00
oscillation start 0

distance 50.00   [crystal to detector distance]
space group P3
mosaicity   0.80
[file sites]
raw data file '../osn33_01.####'
film output file 'osn33_####.x'
peak search file 'peaks.file'

[auto.dat]

peak search file peaks.file
weak level 20

fit x beam y beam cell crystal rotx roty rotz

write predictions
print statistics
print zones

go go go

weak level 1.0   [Reset for refinement]

Refinement

[refinment macro - originally by A. Berghuis, modified by Q. Hoang]
[modifications based on HKL manual]
[modified for refinement of osteo collected at Andy Howard's beamline at APS, Aug 200]

start refinement
resolution limits 40 5.0
fit crystal rotx roty rotz
go go go go go go go go go go
fit cell y beam x beam
go go go go go go go go go go
fix all

resolution limits 40. 4.0
fit crystal rotx roty rotz
go go go go go go go go go go
fix all
fit cell
go go go go go go go go go go
fix all
fit cassette rotx roty rotz
fit crossfire x y xy
go go go go go go go go go go
fix all

resolution limits 40. 3.5

fit crystal rotx roty rotz
go go go go go go go go go go
fit cell y beam x beam
go go go go go go go go go go
fix all

resolution limits 40. 3.0

fit crystal rotx roty rotz
go go go go go go go go go go
fix all

resolution limits 40. 2.3   [change these to approp. res]
fit cell y beam x beam
go go go go go go go go go go

```
fit y scale skew
fit distance
fit cassette rotx roty rotz
fit crossfire x y xy
go go go go go go go go go go
fix all

resolution limits 40. 2.1              [change these to approp. res]
fit cell y beam x beam
go go go go go go go go go go
fit y scale skew
fit distance
fit cassette rotx roty rotz
fit crossfire x y xy
go go go go go go go go go go
fix all

resolution limits 40. 1.80             [change these to approp. res]
fit cell y beam x beam
go go go go go go go go go go
fit y scale skew
fit distance
fit cassette rotx roty rotz
fit crossfire x y xy
go go go go go go go go go go
fix all

calculate go
print profiles                         [default 3x3 grid]
end of pack
```

Scaling and Data reduction

[This is copied from Andy Howard's scl.in, Quyen Aug 2000]

scalepack log file 'scalepack.log'
rejection probability 0.0001
@reject
write rejection file
resolution 2.0
number of zones 10
estimated error
0.05 0.06 0.065 0.07 0.075
0.065 0.065 0.025 0.015 0.01
error scale factor 1
default scale 1
reference film 7
scale restrain 0.01
B restrain 0.1
space group P3121
output file 'osn33.sca'
scale anomalous
ignore overloads
add partials 1 to 100
fit crystal a* 1 to 100
fit crystal c* 1 to 100
fit batch rotx 1 to 100
fit batch roty 1 to 100
fit crystal mosaicity 1 to 100
postrefine 10
write rejection file
format denzo_ip
sector 1 to 100
FILE 1 'osn33_####.x'
;

Appendix 3

Script for phase determination in SOLVE and solvent flattening with

RESOLVE

```
#!/bin/csh
#
#   solve.com file to run SOLVE/RESOLVE version 2.01
#
# set CCP4 and SOLVETMPDIR  and SYMOP variables:
#
setenv CCP4_OPEN UNKNOWN
setenv SOLVETMPDIR /tmp
setenv SYMOP /usr/local/lib/solve/symop.lib
#
#   use all system resources:
#
unlimit
#
#
# command file to solve an SAD dataset (Ca 1-wavelength)
#
#
solve<<EOD > solve.log
logfile solve.logfile

resolution 20 2.0
cell 51.484 51.484 35.387 90.000 90.000 120.000
symfile /usr/local/lib/solve/p3121.sym

readdenzo      ! readformatted/readdenzo/readtrek/readccp4_unmerged
premerged             ! premerged/ unmerged
read_intensities     ! read_intensities/read_amplitudes
fixscattfactors       ! fixscattfactors/refscattfactors

newatomtype Ca
aval 15.634 7.952 0.8537
```

```
bval -0.0074 0.6089 10.3116 25.99
cval -14.875
fprimv 0.2940
fprprv 1.5254

mad_atom Ca

lambda 1
label SAD data for Ca
rawmadfile osn33.sca
#fprprv_mad 5.0
fprprv_mad 1.5254

nsolsite_deriv 7          ! 2 atoms max
SAD                      ! solve an SAD dataset
EOD
#
#  Now do density modification and build a model:
resolve<<EOD>resolve.log
solvent_content 0.40
EOD
#  All done; your output phases are in resolve.mtz and
#   your model is in resolve.pdb
#
```

Appendix 4.1

Fortran source code for the program SURFACE for generating a slab of hydroxyapatite lattice

The output from this program feeds into the program CALDISTANCE/ROTATE

```
C---
C PROGRAM SURFACE
C THIS PROGRAM GENERATE A SLAB OF ATOMS BOUND BY INPUT MILLER INDEXED
PLANE
C A NUMBER OF SLABS ARE BEING OUTPUT REPRESENTING ALL POSSIBLE SURFACE
C REPRESENTATIONS. THE SURFACE CLOSEST FROM THE ORIGIN IS THE REQUESTED
C INPUTS:
C  UNIT 5: 1ST CARD: NAME OF PARAMETER FILE
C          2ND CARD: NAME OF COORDINATE FILE
C          3RD CARD: MILLER INDEX H, K, L
C          4TH CARD: RADIUS AND THICKNESS OF OUTPUT PLANE (XRAD,XTHICK)
C                    ANGSTROM   NO. UNITCELL REPEAT (CAN BE
FRACTIONAL)
C          5TH CARD: 0 FOR PRESENTATION AS ORIGINAL DIRECTION.
C                    1    NORMAL TO X AXIS
C                    2    NORMAL TO Y AXIS
C                    3    NORMAL TO Z AXIS
C
C  UNIT COORDINATE FILE
C          1ST CARD: NATOM (NUMBER OF ATOMS IN ASYMMETRIC UNIT)
C          2: 1+N  : ATOMNAME,X,Y,Z (FRACTION COORDINATES FOR EACH
ATOMS)
C                    (A4,3F10.3)
C  UNIT PARAMETER FILE
C          1ST CARD: LOGFILE NAME (LOGFILE=LOGFILE NAME)
C          2ND CARD: LATTICE TYPE (LATTICE=LATTICE TYPE)
C          3RD CARD: A,B,C,ALPHA,BETA,GAMMA (ANGSTROM,DEGREE)
C          4TH CARD: NSYM (NUMBER OF SYMMETRY CARD)
C          5 TO 4+N: SYMMETRY CARD AS APPEARS IN INTERNATIONAL TABLE OF
CRYST..
C OUTPUT:
C    STARTING FROM UNIT 12,(13,14,...) FOR DIFFERENT UNIQUE PLANES
C  UNIT 12: PDB STYLE COORDINATES:
C    ORIGINAL  NOW         DESCRIPTION
C    ========  ====        ==============
C    ATOMNO    NATM:       NUMBER OF ATOM
```

```
C      ATOMNAME   ATM:      ATOM NAME
C      RESNAME    NSYM:     SYMMETRY OPERATOR
C      CHAIN      SIGNA:    P FOR POSITIVE, N FOR NEGATIVE, TRANSLATION
ALONG A
C      RESNUM     TRANA:    NUMBER OF UNIT CELL TRANSLATION ALONG A
C      X,Y,Z:               CARTESION (ORTHOGONAL) COORDINATES
C      OCC,TEMP TRANS B,C:  NUMBER OF UNIT CELL TRANSLATION ALONG B & C
CATOM    NATM ATM  NSYM SIGNA TRANA X      Y        Z       TRANS B,C
CATOM       1 OH2 12  P   10       -5.02   2.12     3.54    -12. -17.
CATOM       1 O   ASP A   4        -5.260  19.488   16.799  -23.  30.
C
C    THE OUTPUT COORDINATES ARE GIVEN WITH A NEGATIVE DISTANCE FROM THE
C    ORIGIN. ONE CAN REQUEST THE PLANE TO BE ROTATED TO NORMAL TO ONE
OF
C    THE THREE AXES.
C
C    MODIFIED FROM MAKO.F, DAN YANG, WRITTEN IN AUGUEST, 1995
C
C    THE PROGRAM WILL GENERATE UP TO 20000 ATOMS.
C    SYMMETRY OPERATION PART OF THIS PROGRAM WAS TAKEN FROM GREF
WRITTEN BY W. FUREY, V. A.
C    MEDICAL CENTER AND UNIV. OF PITTSBURGH, AUGUST, 1982
C    THIS VERSION IS FULLY GENERAL AND WILL WORK WITH
C    ANY SPACE GROUP, INCLUDING NON-STANDARD SETTINGS.
C
C  DANIEL YANG, DEPARTMENT OF BIOCHEMISTRY, MCMASTER UNIVERSITY,
HAMILTON
C  TEL: 905-525-9140 EXT 22455
C  FAX: 905-522-9033
C  E-MAIL: YANG@MCMASTER.CA
C  JUNE 2000
C
C
       CHARACTER LATT*1,SYMLIN*80
       CHARACTER PNAME*4,FILNAM*80
       LOGICAL IFIRST,IANO,R_LOG
       COMMON /CELL/ PARA(6),UT(3,3),UTINV(3,3)
       COMMON /RCELL/ AS,BS,CS,CAS,CBS,CGS
       COMMON /SYMTRY/ NSYM,FS(3,3,24),TS(3,24),ISS(3,3,24),ITS(3,24)
       DIMENSION SYMLIN(24)
C--- ARRAYS IN ARRAYS XXF,XYF,XZF MUST BE DIMENSIONED FOR AT LEAST THE
C--- TOTAL NUMBER OF ATOMS INPUT (IN ASYMMETRIC UNIT)
       CHARACTER*4 ATMIN(50),ATM1(2000),ATMTMP(2000)
       DIMENSION XXF(50),XYF(50),XZF(50)
       DIMENSION XX1(2000),XY1(2000),XZ1(2000),ISYMM(2000)
       DIMENSION XUNIQD(2000),UNQDST(2000)
       DIMENSION XX1TMP(2000),XY1TMP(2000),XZ1TMP(2000)
       DIMENSION ISYTMP(2000)
C XXF,XYF,XZF,ATMIN: INPUT X,Y,Z FRACTIONAL COORDINATES, ATOM NAME
C XX1,XY1,XZ1,ATM1 : TRANSFORMED COORDINATES, ATOM NAME
C ISYMM            : SYMMETRY OPERATION
C   XM,UXM ORIENTATION MATRICES FOR REQUESTED PLANE
       DIMENSION XM(3,3),UXM(3,3),CARTXM(3,3),ISSS(3,3,24)
C TEMPARORY VAR FOR OUTPUT SIGN BIT FOR X-TRANSLATION
```

```
      CHARACTER XASIGN*1
      PNAME='SURF'
C---
C--- READ NAME OF PARAMETER FILE
      READ(5,1)FILNAM
 1    FORMAT(A)
C---
C--- GET THE VITAL PARAMETERS
      CALL GETPMS(PNAME,FILNAM,A,B,C,AL,BE,GA,LATT,NSYM,SYMLIN,8,9,
     $R_LOG)
C---
C--- GET NAME OF INPUT COORDINATE FILE
      READ(5,1)FILNAM
      WRITE(6,2)FILNAM
      IF(R_LOG)WRITE(9,2)FILNAM
 2    FORMAT(/1X,'INPUT COORDINATE FILE : ',A)
      OPEN(UNIT=11,FILE=FILNAM,FORM='FORMATTED',STATUS='OLD')
C---
      PARA(1)=A
      PARA(2)=B
      PARA(3)=C
      PARA(4)=AL
      PARA(5)=BE
      PARA(6)=GA
      WRITE(6,10)PARA
      IF(R_LOG)WRITE(9,10)PARA
 10   FORMAT(/1X,'A=',F9.4,' B=',F9.4,' C=',F9.4,' AL=',F9.4,
     $' BE=',F9.4,' GA=',F9.4,/)
C---
C--- GET RECIPROCAL CELL CONSTANTS, ORTHOGONALIZATION MATRIX AND ITS
C--- INVERSE

      CALL ORTHOG
C---
C--- GET SYMMETRY INFO
      DO 50 I=1,NSYM
      J=I
      WRITE(6,40)J,SYMLIN(J)
      IF(R_LOG)WRITE(9,40)J,SYMLIN(J)
 40   FORMAT(1X,'EQUIVALENT POSITION',I3,' : ',A)
 50   CALL TRANSL(J,SYMLIN(J),R_LOG)


C***  SORT ISS FOR LATER USAGE ON FOR NSYMI EQUVALENT INDEX
      NSYMI=0
      CALL SORTID(NSYMI,ISSS)

      READ(5,*) IH,IK,IL
C*** GENERATE NSYM EQUVALENT INDEX AND SELECT THE LARGEST ONE
C---
C--- PUT INDICES IN STANDARD REPRESENTATION
C--    I.E. WITH THE MAXIMUM PACKED INDEX, FRIEDEL ASSUMMED
C--
```

```
      CALL STDREF(IH,IK,IL,ISSS,NSYMI)

C** GET CARTESION COORDINATES OF THREE AXIAL VECTORS CORRESPONDING TO
C** THE INPUT MILLER INDEX
C** AND GENERATE EQUATION OF PLANE IN ORTHOGONAL SYSTEM
C**  FIRST CONVERT THE THREE AXIAL VECTORS TO CARTESIAN
C**  FRACTIONAL COORDINATES FOR THREE AXIAL VECTORS ARE
C**  (1,0,0); (0,1,0); (0,0,1)
C**  CONVERTED ORTH COORDINATES  ARE
C**  STORED IN CARTXM AS ROW VECTORS
C**  CONSTRUCT MATRIX
      CARTXM(1,1)=1.
      CARTXM(1,2)=0.
      CARTXM(1,3)=0.
      CALL MATMX(UTINV,CARTXM(1,1),CARTXM(1,2),CARTXM(1,3))
      CARTXM(2,1)=0.
      CARTXM(2,2)=1.
      CARTXM(2,3)=0.
      CALL MATMX(UTINV,CARTXM(2,1),CARTXM(2,2),CARTXM(2,3))
      CARTXM(3,1)=0.
      CARTXM(3,2)=0.
      CARTXM(3,3)=1.
      CALL MATMX(UTINV,CARTXM(3,1),CARTXM(3,2),CARTXM(3,3))
C**  FRACTIONAL COORDINATES FOR THREE AXIAL VECTORS ARE
C**  (XA,0,0); (0,XB,0); (0,0,XC)
C**  CONVERTED ORTH COORDINATES  ARE
C**  (XXA,XYA,XZA); (XXB,XYB,XZB); (XXC,XYC,XZC)
C**  CONSTRUCT MATRIX
      IZERO=0
      XZERO=0.000000000000000001
      ZERO=0.0
      XH=FLOAT(IH)
      XK=FLOAT(IK)
      XL=FLOAT(IL)
      IF(IH.EQ.IZERO) XH=XZERO
      IF(IK.EQ.IZERO) XK=XZERO
      IF(IL.EQ.IZERO) XL=XZERO
      XA=1./XH
      XB=1./XK
      XC=1./XL
      XM(1,1)=XA
      XM(1,2)=0.
      XM(1,3)=0.
      CALL MATMX(UTINV,XM(1,1),XM(1,2),XM(1,3))
      XM(2,1)=0.
      XM(2,2)=XB
      XM(2,3)=0.
      CALL MATMX(UTINV,XM(2,1),XM(2,2),XM(2,3))
      XM(3,1)=0.
      XM(3,2)=0.
      XM(3,3)=XC
      CALL MATMX(UTINV,XM(3,1),XM(3,2),XM(3,3))

C SPECIAL CASES: TO AVOID ILL DEFINED MATRIX
```

```
C  WITH OKL, H0L, HK0, 00L, 0K0, H00

C  OKL
      IF(IH.EQ.0.AND.IK.NE.0.AND.IL.NE.0) THEN
C     GET DIFFERENCE VECTOR BETWEEN B & C INTERCEPT
      XTEM=XM(2,1)-XM(3,1)
      YTEM=XM(2,2)-XM(3,2)
      ZTEM=XM(2,3)-XM(3,3)
C     GET NORMAL TO DIFFERENCE VECTOR AND A
      XTEMP=YTEM*CARTXM(1,3)-ZTEM*CARTXM(1,2)
      YTEMP=ZTEM*CARTXM(1,1)-XTEM*CARTXM(1,3)
      ZTEMP=XTEM*CARTXM(1,2)-YTEM*CARTXM(1,1)

C  H0L
      ELSEIF (IH.NE.0.AND.IK.EQ.0.AND.IL.NE.0) THEN
C     GET DIFFERENCE VECTOR BETWEEN B & C INTERCEPT
      XTEM=XM(1,1)-XM(3,1)
      YTEM=XM(1,2)-XM(3,2)
      ZTEM=XM(1,3)-XM(3,3)
C     GET NORMAL TO DIFFERENCE VECTOR AND B
      XTEMP=YTEM*CARTXM(2,3)-ZTEM*CARTXM(2,2)
      YTEMP=ZTEM*CARTXM(2,1)-XTEM*CARTXM(2,3)
      ZTEMP=XTEM*CARTXM(2,2)-YTEM*CARTXM(2,1)

C  HK0
      ELSEIF (IH.NE.0.AND.IK.NE.0.AND.IL.EQ.0) THEN
C     GET DIFFERENCE VECTOR BETWEEN B & C INTERCEPT
      XTEM=XM(1,1)-XM(2,1)
      YTEM=XM(1,2)-XM(2,2)
      ZTEM=XM(1,3)-XM(2,3)
C     GET NORMAL TO DIFFERENCE VECTOR AND C
      XTEMP=YTEM*CARTXM(3,3)-ZTEM*CARTXM(3,2)
      YTEMP=ZTEM*CARTXM(3,1)-XTEM*CARTXM(3,3)
      ZTEMP=XTEM*CARTXM(3,2)-YTEM*CARTXM(3,1)

C  00L
      ELSEIF (IH.EQ.0.AND.IK.EQ.0.AND.IL.NE.0) THEN
C   IE. NORMAL TO A & B AXES, A X B
C
      XTEMP=CARTXM(1,2)*CARTXM(2,3)-CARTXM(2,2)*CARTXM(1,3)
      YTEMP=CARTXM(1,3)*CARTXM(2,1)-CARTXM(2,3)*CARTXM(1,1)
      ZTEMP=CARTXM(1,1)*CARTXM(2,2)-CARTXM(2,1)*CARTXM(1,2)

C  0K0
      ELSEIF (IH.EQ.0.AND.IK.NE.0.AND.IL.EQ.0) THEN
C   IE. NORMAL TO A & C AXES, A X C CROSSED PRODUCT
C
      XTEMP=CARTXM(1,2)*CARTXM(3,3)-CARTXM(3,2)*CARTXM(1,3)
      YTEMP=CARTXM(1,3)*CARTXM(3,1)-CARTXM(3,3)*CARTXM(1,1)
      ZTEMP=CARTXM(1,1)*CARTXM(3,2)-CARTXM(3,1)*CARTXM(1,2)

C  H00
```

```
C   IE. NORMAL TO B & C AXES, B X C
C
       ELSEIF (IH.NE.0.AND.IK.EQ.0.AND.IL.EQ.0) THEN
       XTEMP=CARTXM(2,2)*CARTXM(3,3)-CARTXM(3,2)*CARTXM(2,3)
       YTEMP=CARTXM(2,3)*CARTXM(3,1)-CARTXM(3,3)*CARTXM(2,1)
       ZTEMP=CARTXM(2,1)*CARTXM(3,2)-CARTXM(3,1)*CARTXM(2,2)


C  HKL GENERAL REFLECTIONS
       ELSE

C**    INVERT MATRIX XM GIVE DIRECTION COSINE FOR PLANE
       CALL UINV(XM,UXM)
       XTEMP=UXM(1,1)+UXM(1,2)+UXM(1,3)
       YTEMP=UXM(2,1)+UXM(2,2)+UXM(2,3)
       ZTEMP=UXM(3,1)+UXM(3,2)+UXM(3,3)
       ENDIF

C** NORMALIZE X,Y,Z COSINES
       XCOSIN=XTEMP/SQRT(XTEMP**2+YTEMP**2+ZTEMP**2)
       YCOSIN=YTEMP/SQRT(XTEMP**2+YTEMP**2+ZTEMP**2)
       ZCOSIN=ZTEMP/SQRT(XTEMP**2+YTEMP**2+ZTEMP**2)
C* CHANGE TO POSITIVE DIRECTION
       IF(ZCOSIN.LT.0.0) THEN
       XCOSIN=-XCOSIN
       YCOSIN=-YCOSIN
       ZCOSIN=-ZCOSIN
       ENDIF
       IF(YCOSIN.LT.0.0) THEN
       XCOSIN=-XCOSIN
       YCOSIN=-YCOSIN
       ZCOSIN=-ZCOSIN
       ENDIF
       IF(XCOSIN.LT.0.0) THEN
       XCOSIN=-XCOSIN
       YCOSIN=-YCOSIN
       ZCOSIN=-ZCOSIN
       ENDIF

       READ(5,*) XRAD,XTHICK
       READ(5,*) NORIEN

C READ FRACTIONAL COORDINATES OF A COMPLETE ASYMMETRIC UNIT
C      OF ATOMS
C
       READ(11,*) NATOM
       WRITE(6,1005)
 1005 FORMAT(' INPUT FRACTIONAL COORDINATES')
       DO 101, I=1,NATOM
       READ(11,102,END=101)ATMIN(I),XXF(I),XYF(I),XZF(I)
       WRITE(6,102) ATMIN(I),XXF(I),XYF(I),XZF(I)
 102   FORMAT(A4,3F10.3)
 101   CONTINUE
```

```
C** GENERATE A COMPLETE UNIT CELL BOX OF ATOMS
C**        (IN THE FIRST ALL POSITIVE UNIT CELL)
      IATOM1=0
      DO 201 I=1,NATOM
      DO 202 ISYM=1,NSYM
      XTEMP=XXF(I)
      YTEMP=XYF(I)
      ZTEMP=XZF(I)
      IATOM1=IATOM1+1
C APPLY SYMMETRY OPERATION IN FRACTIONAL COORDINATES
C  AND USE A MOD FUNCTION TO PUT EVERYTHING IN THE FIRST
C    POSITIVE UNIT CELL
C    OUTPUT COORDINATE: XTEMP,YTEMP,ZTEMP
      CALL ROTRAN(ISS(1,1,ISYM),ITS(1,ISYM),XTEMP,YTEMP,ZTEMP)
C SAVE COORDINATES TO ARRAY XX1,XY1,XZ1 (IN FRACTIONAL FORM)
      XX1TMP(IATOM1)=XTEMP
      XY1TMP(IATOM1)=YTEMP
      XZ1TMP(IATOM1)=ZTEMP
      ATMTMP(IATOM1)=ATMIN(I)
      ISYTMP(IATOM1)=ISYM
  202 CONTINUE
  201 CONTINUE

C     WRITE(6,701)
  701 FORMAT(' SYMMETRY EQUALENT POSITIONS')
      WRITE(6,702)(ATMTMP(I),XX1TMP(I),XY1TMP(I)
     $,XZ1TMP(I),I=1,IATOM1)
  702 FORMAT(3X,A4,4X,3F10.3)
C***  SORT ATOM ARRAY TO REMOVE REDUNDANT ATOMS AT SPECIAL POSITION
C***  XYZ POSITION PACKED BY COVERTING TO INTEGER AND PACKED
C***   INTO ONE WORD, ROUNDING UP WITH TWO SIGNIFICANT FIGURES SHOULD
C      BE ACCURATE ENOUGH ??
      IU=1
      XX1(1)=XX1TMP(1)
      XY1(1)=XY1TMP(1)
      XZ1(1)=XZ1TMP(1)
      ATM1(1)=ATMTMP(1)
      ISYMM(1)=ISYTMP(1)
      DO 601 I=2,IATOM1
      IXYZPK=1000000*IFIX(100.*(XX1TMP(I)+0.005))
     $+        1000*IFIX(100.*(XY1TMP(I)+0.005))
     $+             IFIX(100.*(XZ1TMP(I)+0.005))
      DO 602 J=1,IU
      IXYZ = 1000000*IFIX(100.*(XX1(J)+0.005))
     $+        1000*IFIX(100.*(XY1(J)+0.005))
     $+             IFIX(100.*(XZ1(J)+0.005))
      IF (IXYZ.EQ.IXYZPK) GOTO 601
  602 CONTINUE
      IU=IU+1
      XX1(IU)=XX1TMP(I)
      XY1(IU)=XY1TMP(I)
      XZ1(IU)=XZ1TMP(I)
      ATM1(IU)=ATMTMP(I)
      ISYMM(IU)=ISYTMP(I)
```

```
  601   CONTINUE
C        WRITE(6,1006)
C 1006 FORMAT(' COMPLETE LIST OF ATOMS IN UNIT CELL:')
C        WRITE(6,1002)(ATMTMP(I),XX1TMP(I),XY1TMP(I)
C       $,XZ1TMP(I),I=1,IATOM1)
 1002 FORMAT(2X,A8,3F10.3)
      IATOM1=IU
      WRITE(6,703)IU
  703  FORMAT(' NO. OF UNIQUE ATOMS :',I5)
      WRITE(6,702)(ATM1(I),XX1(I),XY1(I),XZ1(I),I=1,IU)
      WRITE(6,1004)
 1004 FORMAT('    ')

C**
C** CALCULATE AND STORE DISTANCES OF ATOM IN ONE UNIT CELL
C
      DO 401 I=1,IATOM1
      XTEMP=XX1(I)
      YTEMP=XY1(I)
      ZTEMP=XZ1(I)
C** CONVERT ATOMS TO ORTHOGONAL SYSTEM
      CALL MATMX(UTINV,XTEMP,YTEMP,ZTEMP)
C** GET DOT PRODUCT WITH DIRECTION COSINE
C**  AND STORE AS DISTANCE FROM PLANE
      XUNIQD(I)=XTEMP*XCOSIN+YTEMP*YCOSIN+ZTEMP*ZCOSIN
  401   CONTINUE

C**   NOW SORT XUNIQD AND STORE IN ARRAY UNQDST
C**     WITH NUNIQ = NO. OF UNIQUE POSITIONS
      CALL SORTX(XUNIQD,IATOM1)
      UNQDST(1)=XUNIQD(1)
      NUNIQ=1
      DO 402 I=2,IATOM1
      IF(ABS(UNQDST(NUNIQ)-XUNIQD(I)).LT.0.0001) GOTO 402
      NUNIQ=NUNIQ+1
      UNQDST(NUNIQ)=XUNIQD(I)
  402   CONTINUE
      WRITE(6,1010)NUNIQ
 1010 FORMAT(' NUMBER OF UNIQUE PLANES : ', I5)
      WRITE(6,1011)(UNQDST(I),I=1,NUNIQ)
 1011 FORMAT('  UNIQUIE DISTANCES : ', 5F10.3)

C*    CALCULATE DISTANCE TO THE APEX OF UNIT CELL
C     XAPEX, YAPEX, ZAPEX
      XAPEX=1.
      YAPEX=1.
      ZAPEX=1.
      CALL MATMX(UTINV,XAPEX,YAPEX,ZAPEX)
      XTHICK=XTHICK*(XAPEX*XCOSIN+YAPEX*YCOSIN+ZAPEX*ZCOSIN)
C* HERE WE START TO GENERATE COORDINATES FOR OUTPUT
C*  A NUMBER OF FILES WILL BE GENERATED STARTING WITH
C*   FORT.12, EACH OF A UNIQUE SET OF ATOMS ON THE SURFACE
      NOUT=11
      DO 501 IUNIQ=1,NUNIQ
```

```
C
      XZERO1=  0.0
      XZERO = - (0.0+UNQDST(IUNIQ)     - 0.001)
      XCUT  = - (XTHICK+UNQDST(IUNIQ) - 0.001)
      NOUT  = NOUT+1
      WRITE (NOUT,1013)XZERO1,XZERO1,XZERO1,XZERO,XZERO
 1013 FORMAT('ATOM','   9999',2x,'ORI ',3X,1X,1X,'9999'
     $,F12.3,2F8.3,2F6.2)
C**
C** SEARCH A RETANGULAR BOX WITH XRADSK OF ATOMS IN ORIGINAL
C**    COORDINATE SYSTEM
C     ITA, ITB AND ITC ARE INDICES FOR THE NESTED LOOP
C     XRAD IS EXTENDED 2. FOR POSSIBLE SKEWED CELL
C
      XRADSK=XRAD*2.
      ITA=IFIX(XRADSK/PARA(1)+0.5)
      ITB=IFIX(XRADSK/PARA(2)+0.5)
      ITC=IFIX(XRADSK/PARA(3)+0.5)
      IAA=0
      OCC=1.0
      TEMP=20.
      DO 301 IA=-ITA,ITA
      IABSO=IABS(IA)
      XASIGN='P'
      IF(IA.LT.0) XASIGN='N'
      DO 302 IB=-ITB,ITB
      XIB=FLOAT(IB)
      DO 303 IC=-ITC,ITC
      XIC=FLOAT(IC)
      DO 304 IATOM=1,IATOM1
      XTEMP=XX1(IATOM)+FLOAT(IA)
      YTEMP=XY1(IATOM)+FLOAT(IB)
      ZTEMP=XZ1(IATOM)+FLOAT(IC)

C** CONVERT ATOMS TO ORTHOGONAL SYSTEM
      CALL MATMX(UTINV,XTEMP,YTEMP,ZTEMP)

C** GENERATE SLAB IN ORTHOGONAL SYSTEM
C* BY USING DISTANCE CUTOFF OF XRAD AND XTHICK
C*
C*  FIRST: GET THE PROJECTION ON DIRECTION COSINE BY
C*         COMPUTING THE DOT PRODUCT
C*            THEN NORMALIZED DIRECTION COSINE (XCOSIN..
C*
      XDOT=XTEMP*XCOSIN+YTEMP*YCOSIN+ZTEMP*ZCOSIN
      XPROJ=XDOT*XCOSIN
      YPROJ=XDOT*YCOSIN
      ZPROJ=XDOT*ZCOSIN
      XDISK=(XTEMP-XPROJ)**2+(YTEMP-YPROJ)**2+(ZTEMP-ZPROJ)**2
      XDISK=SQRT(XDISK)

      IF(XDISK.GT.XRAD) GOTO 304
C*    SECOND CHECK DISTANCE FROM ORIGIN
      XDISK=XTEMP*XCOSIN+YTEMP*YCOSIN+ZTEMP*ZCOSIN
```

```
C*   GET MINIMUM DISPLACEMENT FROM ORIGIN TO GENERATE
C*      DIFFERENT POSSIBLE SURFACES
        IF(XDISK.GT.XZERO) GOTO 304
        IF(XDISK.LT.XCUT) GOTO 304


C **
C ** ROTATE PLANE ACCORDING TO NORIEN
        IF (NORIEN.EQ.0) GOTO 305
        IF (NORIEN.EQ.1) THEN
C   ROTATE TO BE NORMAL TO X
C    FIRST ROTATE Z WITH ATAN(Y/X)
C     THEN  ROTATE Y WITH ASIN(Z)
        THE1=0.
        IF (ABS(XCOSIN).GT.0.00001)THE1=ATAN2(YCOSIN,XCOSIN)
        THE2=-ASIN(ZCOSIN)
        CALL ROTZ(XTEMP,YTEMP,ZTEMP,THE1)
        CALL ROTY(XTEMP,YTEMP,ZTEMP,THE2)


        ELSEIF (NORIEN.EQ.2) THEN
C   ROTATE TO BE NORMAL TO Y
C    FIRST ROTATE Z WITH ATAN(X/Y)
C      THEN ROTATE X WITH ASIN(Z)
        THE1=0.0
        IF(ABS(YCOSIN).GT.0.00001)THE1=-ATAN2(XCOSIN,YCOSIN)
        THE2=ASIN(ZCOSIN)
        CALL ROTZ(XTEMP,YTEMP,ZTEMP,THE1)
        CALL ROTX(XTEMP,YTEMP,ZTEMP,THE2)
        ELSE
C   ROTATE TO BE NORMAL TO Z
C    FIRST ROTATE Z WITH ATAN(X/Y)
C      THEN ROTATE X WITH -ACOS(Z)
        THE1=0.0
        IF(ABS(YCOSIN).GT.0.00001)THE1=-ATAN2(XCOSIN,YCOSIN)
        THE2=-ACOS(ZCOSIN)
        CALL ROTZ(XTEMP,YTEMP,ZTEMP,THE1)
        CALL ROTX(XTEMP,YTEMP,ZTEMP,THE2)
        ENDIF



C OUTPUT: (NOUT)
C   UNIT 12: PDB STYLE COORDINATES:
C      ORIGINAL   NOW          DESCRIPTION
C      ========   ====         ==============
C      ATOMNO     NATM:        NUMBER OF ATOM
C      ATOMNAME   ATM:         ATOM NAME
C      RESNAME    NSYM:        SYMMETRY OPERATOR
C      CHAIN      SIGNA:       P FOR POSITIVE, N FOR NEGATIVE FOR
TRANSLATION ALONG A
C      RESNUM     TRANA:       NUMBER OF UNIT CELL TRANSLATION ALONG A
C      X,Y,Z:                  CARTESION (ORTHOGONAL) COORDINATES
C      OCC,TEMP TRANS B,C:  NUMBER OF UNIT CELL TRANSLATION ALONG B & C
CATOM    NATM ATM  NSYM SIGNA TRANA X      Y        Z       TRANS B,C
```

```
CATOM        1  OH2 12  P  10       -5.02    2.12    3.54   -12. -17.
CATOM        1  O   ASP 4   4       -5.260 19.488  16.799  -23.  30.
C     THE OUTPUT COORDINATES ARE GIVEN WITH A NEGATIVE DISTANCE FROM THE
C     ORIGIN. ONE CAN REQUEST THE PLANE TO BE ROTATED TO NORMAL TO ONE
OF
C     THE THREE AXES.
 305  IAA=IAA+1
      WRITE (NOUT,1003)IAA,ATM1(IATOM),ISYMM(IATOM),XASIGN,IABSO,
     $XTEMP,YTEMP,ZTEMP,XIB,XIC
1003  FORMAT('ATOM',I7,2X,A4,I3,1X,A1,I4,F12.3,2F8.3,2F6.2)
 304  CONTINUE
 303  CONTINUE
 302  CONTINUE
 301  CONTINUE

 501  CONTINUE
      STOP
      END

      SUBROUTINE GETPMS(STRNG,FILNAM,A,B,C,AL,BE,GA,LATT,NSYM,SYMLIN,
     $LP,LG,R_LOG)
      CHARACTER STRNG*(*),LATT*1,FILNAM*80,TIM*8,DAT*9
      CHARACTER LINE*80,SYMLIN*80
      LOGICAL R_LOG
      DIMENSION SYMLIN(24)
C--- ROUTINE TO READ THE MASTER PARAMETER FILE, PICK UP CELL CONSTANTS
C--- AND SYMMETRY INFO, LATTICE TYPE, AND IF REQUESTED, OPEN RUNNING
LOG FILE.
C--- IT ALSO PLACES A TIME STAMP IN THE RUNNING LOG LOG INDICATING WHAT
PROGRAM
C--- IS RUNNING AND WHEN IT STARTED.
C---
C--- GET NAME OF RUNNING LOG FILE (FIRST ENTRY IN PARAM FILE), IF NULL
C--- SET FLAG ACCORDINGLY
      OPEN(UNIT=LP,FILE=FILNAM,FORM='FORMATTED',STATUS='OLD')
      READ(LP,10)LINE
 10   FORMAT(A)
C---
C--- DO WE WANT A RUNNING LOG FILE?
      I=INDEX(LINE,'LOGFILE=')
      IF(I.EQ.0)I=INDEX(LINE,'LOGFILE=')
      IF((I.EQ.0).OR.(LINE(I+8:I+11).EQ.'NULL').OR.(LINE(I+8:I+11).EQ.
     $'NULL'))THEN
C--- NO RUNNING LOG FILE WANTED
      R_LOG=.FALSE.
C---
      ELSE
C---
C--- OPEN RUNNING LOG FILE
      R_LOG=.TRUE.
      OPEN(UNIT=LG,FILE=LINE(I+8:80),FORM='FORMATTED',STATUS='UNKNOWN',
     $ACCESS='APPEND')
C---
C---PLACE TIME STAMP IN THE LOG FILE
```

```
        CALL TIME(TIM)
        CALL DATE(DAT)
        WRITE(LG,20)STRNG,DAT,TIM
 20     FORMAT(//,' ************** ENTERING ',2X,A,2X,A,2X,A,2X,
      $'**********',//)
        ENDIF
C---
C--- GET LATTICE TYPE
        READ(LP,10)LINE
        I=INDEX(LINE,'LATTICE=')
        IF(I.EQ.0)I=INDEX(LINE,'LATTICE=')
        LATT=LINE(I+8:I+8)
C---
C--- GET CELL CONSTANTS
        READ(LP,*)A,B,C,AL,BE,GA
C---
C--- GET NUMBER OF EQUIVALENT POSITIONS
        READ(LP,*)NSYM
C---
C--- GET THE SYMMETRY OPERATORS
        DO 30 I=1,NSYM
 30     READ(LP,10)SYMLIN(I)
C---
C--- THATS ALL
        CLOSE(UNIT=LP,STATUS='KEEP')
C---
        RETURN
        END

        SUBROUTINE UINV(AA,BB)
C--- ROUTINE TO INVERT A 3X3 MATRIX
C--- MATRIX BB IS INVERSE OF INPUT MATRIX AA
        DIMENSION AA(3,3),BB(3,3)
        BB(1,1)=AA(2,2)*AA(3,3)-AA(3,2)*AA(2,3)
        BB(2,1)=AA(2,3)*AA(3,1)-AA(2,1)*AA(3,3)
        BB(3,1)=AA(2,1)*AA(3,2)-AA(2,2)*AA(3,1)
        BB(1,2)=AA(1,3)*AA(3,2)-AA(1,2)*AA(3,3)
        BB(2,2)=AA(1,1)*AA(3,3)-AA(1,3)*AA(3,1)
        BB(3,2)=AA(1,2)*AA(3,1)-AA(1,1)*AA(3,2)
        BB(1,3)=AA(1,2)*AA(2,3)-AA(1,3)*AA(2,2)
        BB(2,3)=AA(1,3)*AA(2,1)-AA(1,1)*AA(2,3)
        BB(3,3)=AA(1,1)*AA(2,2)-AA(1,2)*AA(2,1)
        DET=AA(1,1)*AA(2,2)*AA(3,3)+AA(2,1)*AA(3,2)*AA(1,3)+
      $AA(3,1)*AA(1,2)*AA(2,3)-AA(3,3)*AA(1,2)*AA(2,1)-
      $AA(1,1)*AA(3,2)*AA(2,3)-AA(3,1)*AA(2,2)*AA(1,3)
        DO 1 I=1,3
        DO 1 J=1,3
 1      BB(J,I)=BB(J,I)/DET
        RETURN
        END

        SUBROUTINE TRANSL(J,LINE,R_LOG)
C--- ROUTINE TO TRANSLATE FREE FORMAT SYMMETRY CARDS INTO ROTATION
C--- MATRICES AND TRANSLATION VECTORS
```

```
      CHARACTER LINE(80),ICH(20)
      LOGICAL R_LOG
      COMMON /SYMTRY/ NSYM,FS(3,3,24),TS(3,24),ISS(3,3,24),ITS(3,24)
      DATA ICH /'1','2','3','4','5','6','7','8','9','X','x','Y','y',
     $'Z','z','+','-',',','/',' '/
      IGN=1
      I=1
      IDIV=0
      DO 10 IJ=1,3
      DO 20 IJK=1,3
      ISS(IJK,IJ,J)=0
   20 FS(IJK,IJ,J)=0.0
      ITS(IJ,J)=0
   10 TS(IJ,J)=0.
      DO 50 ICAR=1,80
      IF(LINE(ICAR).EQ.ICH(20))GO TO 50
      DO 5 NUM=1,19
      IF(ICH(NUM).EQ.LINE(ICAR))GO TO 6
    5 CONTINUE
      WRITE(6,3)LINE(ICAR)
      IF(R_LOG)WRITE(9,3)LINE(ICAR)
    3 FORMAT(/1X,A1,' IS AN INVALID CHARACTER')
      STOP
    6 IF(NUM.EQ.18)IGN=1
      IF(NUM.EQ.18)I=I+1
      IF(NUM.GT.9)GO TO 12
      IF(IDIV)7,7,8
    7 TS(I,J)=FLOAT(NUM)
      IGN=1
      GO TO 50
    8 TS(I,J)=TS(I,J)/FLOAT(NUM)
      IDIV=0
      GO TO 50
   12 IF(NUM.EQ.16)IGN=1
      IF(NUM.GT.17)GO TO 40
      IF(NUM.EQ.17)IGN=-1
      IF(NUM.EQ.17)GO TO 50
      IF((NUM.EQ.10).OR.(NUM.EQ.11))THEN
      FS(1,I,J)=FLOAT(IGN*1)
      ISS(I,1,J)=IGN
      ENDIF
      IF((NUM.EQ.12).OR.(NUM.EQ.13))THEN
      FS(2,I,J)=FLOAT(IGN*1)
      ISS(I,2,J)=IGN
      ENDIF
      IF((NUM.EQ.14).OR.(NUM.EQ.15))THEN
      FS(3,I,J)=FLOAT(IGN*1)
      ISS(I,3,J)=IGN
      ENDIF
   40 IF(NUM.EQ.19)IDIV=1
   50 CONTINUE
C--- MULTIPLY ALL ELEMENTS BY EQUIVALENT OF TWOPI FOR STRUCTURE FACTOR
C--- CALCULATION
      DO 70 I=1,3
```

```
        DO 60 K=1,3
 60     FS(K,I,J)=FS(K,I,J)*2048.
C--- MULTIPLY TRANSLATION TERM BY 24 FOR USE IN ROUTINE STDREF, AND
C--- EQUIVALENT OF TWOPI FOR STRUCTURE FACTOR CALCULATION
        ITS(I,J)=(TS(I,J)*24. + SIGN(.5,TS(I,J)))
 70     TS(I,J)=TS(I,J)*2048.
        RETURN
        END


C-- SUBROUTINE TO APPLY ROTATION AND TRANSLATION
C -- TO FILL IN A UNIT CELL OF ATOMS WITH X,Y,Z ALL POSITIVE
C      IN FRACTIONAL COORDINATE
C-- USING INTEGER MATRIX FROM SUBROUTINE TRANSL

        SUBROUTINE ROTRAN (ISSS,ITSS,XXX,YYY,ZZZ)
        DIMENSION ISSS(3,3),ITSS(3),X(3),SSS(3,3),TSS(3)
        DO 1 I=1,3
        DO 2 J=1,3
        SSS(I,J)=FLOAT(ISSS(I,J))
 2      CONTINUE
        TSS(I)=FLOAT(ITSS(I))/24.
 1      CONTINUE
        X(1)=AMOD((99.+XXX*SSS(1,1)+YYY*SSS(1,2)+ZZZ*SSS(1,3)+TSS(1)),1.)
        X(2)=AMOD((99.+XXX*SSS(2,1)+YYY*SSS(2,2)+ZZZ*SSS(2,3)+TSS(2)),1.)
        X(3)=AMOD((99.+XXX*SSS(3,1)+YYY*SSS(3,2)+ZZZ*SSS(3,3)+TSS(3)),1.)
        XXX=X(1)
        YYY=X(2)
        ZZZ=X(3)
        RETURN
        END

        SUBROUTINE MATMX (AA,X,Y,Z)
        DIMENSION AA(3,3)
        XX=AA(1,1)*X+AA(1,2)*Y+AA(1,3)*Z
        YY=AA(2,1)*X+AA(2,2)*Y+AA(2,3)*Z
        ZZ=AA(3,1)*X+AA(3,2)*Y+AA(3,3)*Z
        X=XX
        Y=YY
        Z=ZZ
        RETURN
        END

        SUBROUTINE ORTHOG
        COMMON /CELL/ PARA(6),UT(3,3),UTINV(3,3)
        COMMON /RCELL/ AS,BS,CS,COSALS,COSBES,COSGAS
C---
C--- COMPUTE RECIPROCAL CELL CONSTANTS FROM DIRECT CELL CONSTANTS
        DEGTOR=ACOS(-1.)/180.
        AL=PARA(4)*DEGTOR
        BE=PARA(5)*DEGTOR
        GA=PARA(6)*DEGTOR
        VOL=PARA(1)*PARA(2)*PARA(3)*SQRT(1.-COS(AL)**2-COS(BE)**2
        $-COS(GA)**2+2.*(COS(AL)*COS(BE)*COS(GA)))
```

```
      AS=(PARA(2)*PARA(3)*SIN(AL))/VOL
      BS=(PARA(1)*PARA(3)*SIN(BE))/VOL
      CS=(PARA(1)*PARA(2)*SIN(GA))/VOL
      COSALS=(COS(BE)*COS(GA)-COS(AL))/(SIN(BE)*SIN(GA))
      COSBES=(COS(GA)*COS(AL)-COS(BE))/(SIN(GA)*SIN(AL))
      COSGAS=(COS(AL)*COS(BE)-COS(GA))/(SIN(AL)*SIN(BE))
C---
C--- LET U BE A 3X3 TRANSFORMATION MATRIX WHICH RELATES A NEW CELL
C--- WITH ORTHOGONAL AXES OF UNIT DIMENSIONS TO THE ORIGINAL UNIT
C--- CELL (CRYSTAL SYSTEM).  THE NEW CELL HAS NEW B PARALLEL TO OLD
C--- B, NEW A PARALLEL TO OLD B CROSS OLD C, AND NEW C PARALLEL TO
C--- NEW A CROSS NEW B.  THEN FRACTIONAL COORDINATES IN THE CRYSTAL
C--- SYSTEM ARE RELATED TO COORDINATES IN THE ORTHOGONAL SYSTEM BY
C--- X,Y,Z (CRYST) = U TRANSPOSE TIMES X,Y,Z (ORTHOGONAL).
C--- INVERSE OF U TRANSPOSE CONVERTS FRACTIONAL CRYSTAL COORDINATES
C--- TO ORTHOGONAL COORDINATES IN ANGSTROMS.
C---
C--- SET UP MATRIX UT (TRANSPOSE OF MATRIX U)
      UT(1,1)=AS
      UT(1,2)=0.
      UT(1,3)=0.
      UT(2,1)=BS*COSGAS
      UT(2,2)=1./PARA(2)
      UT(2,3)=-(1./TAN(AL))/PARA(2)
      UT(3,1)=CS*COSBES
      UT(3,2)=0.
      UT(3,3)=1./(PARA(3)*SIN(AL))
C---
C--- GET INVERSE OF MATRIX UT
      CALL UINV(UT,UTINV)
C---
      RETURN
      END


      SUBROUTINE STDREF(IH,IK,IL,ISSS,NSYMI)
      DIMENSION ISSS(3,3,24)
      DIMENSION IND(3),JH(3)
C--- ROUTINE TO FIND STANDARD REFLECTION (MAX VALUE OF PACKED INDICES),
C--- MULTIPLICITY. MULT= # TIMES REFLECTION OCCURS IN THE
C--- SPHERE. E
C--- THE ROUTINE IS APPLICABLE TO ALL SPACE GROUPS, INCLUDING ANY
C--- NONSTANDARD SETTINGS.
C---    TAKEN FROM PHASES PACKAGE- CMBISO, WITH MODIFICATION
C---  DAN YANG, MCMASTER UNIVERSITY, JUNE, 2000
      IND(1)=IH
      IND(2)=IK
      IND(3)=IL
C-- GENERATE ALL EQUVALENT REFLECTIONS
C--
      WRITE (6,1002)IND
 1002 FORMAT(' INPUT INDEX: ', 3I7,/,'EQUIVALENT INDEX:')
      DO 180 J=1,NSYMI
      DO 110 I=1,3
      JH(I)=0
```

```
        DO 110 K=1,3
110  JH(I)=JH(I)+ISSS(K,I,J)*IND(K)
        WRITE(6,1001)JH
1001 FORMAT(5x,3I20)
180   CONTINUE


C-- GENERATE UNIQUE REFLECTION WITH MAXIMUM PACKED INDEX
        DO 80 J=1,NSYMI
        DO 10 I=1,3
        JH(I)=0
        DO 10 K=1,3
10   JH(I)=JH(I)+ISSS(K,I,J)*IND(K)
        KIND=(JH(1)+512)*1048576+(JH(2)+512)*1024+JH(3)+512
C--- INSURE THAT FIRST NON-ZERO INDEX IS POSITIVE (HEMISPHERE
C--- WITH H .GE. 0)
        ISG=1
        IF(JH(1))40,20,70
20   IF(JH(2))50,30,70
30   IF(JH(3))60,70,70
40   JH(1)=-JH(1)
50   JH(2)=-JH(2)
60   JH(3)=-JH(3)
        ISG=-1
        KIND=(JH(1)+512)*1048576+(JH(2)+512)*1024+JH(3)+512
70   IF(KIND.LE.IMAX)GO TO 80
        IMAX=KIND
80   CONTINUE
        IH=IMAX/1048576
        IKL=IMAX-IH*1048576
        IH=IH-512
        IK=IKL/1024
        IL=IKL-IK*1024-512
        IK=IK-512
        WRITE(6,1003)IH,IK,IL
1003 FORMAT(' INDEX USED: ', 3I7)
        RETURN
        END

        SUBROUTINE SORTX(X,N)
        DIMENSION X(N)
        XMAX=-100000.
        DO 2 J=1,N
        DO 1 I=J,N
        IF(X(I).LE.X(J)) GOTO 1
        XTEMP=X(I)
        X(I)=X(J)
        X(J)=XTEMP
1    CONTINUE
2    CONTINUE
        RETURN
        END

C
```

```
C    ROTATION MATRIX:   COUNTER CLOCKWISE POSITIVE
C         ROT X                 ROT Y              ROT Z
C    (1    0     0 )    (COS   0  SIN)   (COS   -SIN  0)
C    (0    COS -SIN)    ( 0    1   0 )   (SIN    COS  0)
C    (0    SIN +COS)    (-SIN  0  COS)   ( 0      0   1)
C  TO BE CONSISTENT WITH XPLOR, ANGLES ARE NEGATED
C  SO THE RESULT IS CLOCKWISE POSITIVE
       SUBROUTINE ROTX(X,Y,Z,THETA)
       COSTHE=COS(-THETA)
       SINTHE=SIN(-THETA)
       X2=X
       Y2=Y
       Z2=Z
       X=X2
       Y=Y2*COSTHE-Z2*SINTHE
       Z=Y2*SINTHE+Z2*COSTHE
       RETURN
       END
       SUBROUTINE ROTY(X,Y,Z,THETA)
       COSTHE=COS(-THETA)
       SINTHE=SIN(-THETA)
       X2=X
       Y2=Y
       Z2=Z
       Y=Y2
       Z=Z2*COSTHE-X2*SINTHE
       X=Z2*SINTHE+X2*COSTHE
       RETURN
       END
       SUBROUTINE ROTZ(X,Y,Z,THETA)
       COSTHE=COS(-THETA)
       SINTHE=SIN(-THETA)
       X2=X
       Y2=Y
       Z2=Z
       Z=Z2
       X=X2*COSTHE-Y2*SINTHE
       Y=X2*SINTHE+Y2*COSTHE
       RETURN
       END

       SUBROUTINE SORTID(NSYMI,ISSS)
       COMMON /SYMTRY/ NSYM,FS(3,3,24),TS(3,24),ISS(3,3,24),ITS(3,24)
       DIMENSION ISSS(3,3,24)
       NSYMI=1
       DO 1 I=1,3
       DO 1 J=1,3
       ISSS(I,J,1)=ISS(I,J,1)
1      CONTINUE
       DO 4 K1=2,NSYM
       DO 2 K2=1,NSYMI
       IF(ISSS(1,1,K2).EQ.ISS(1,1,K1).AND.
     $    ISSS(1,2,K2).EQ.ISS(1,2,K1).AND.
     $    ISSS(2,1,K2).EQ.ISS(2,1,K1).AND.
```

```
    $   ISSS(2,2,K2).EQ.ISS(2,2,K1))   GOTO 4
2       CONTINUE
        NSYMI=NSYMI+1
        DO 3 I=1,3
        DO 3 J=1,3
        ISSS(I,J,NSYMI)=ISS(I,J,K1)
3       CONTINUE
4       CONTINUE
        WRITE(6,12)
12      FORMAT(' ISS BEFORE SORTING')
        WRITE(6,10)(((ISS(I,J,K),I=1,3),J=1,3),K=1,NSYM)
        WRITE(6,11)
11      FORMAT(' SORTED ISS MATRIX')
        WRITE(6,10)(((ISSS(I,J,K),I=1,3),J=1,3),K=1,NSYMI)
10      FORMAT(3I10)
        RETURN
        END
```

Appendix 4.2

Fortran source code for the program CAL_CEN_MOLE generate the two

unique calcium positions in hydroxyapatite

The output from this program feeds into the program CALDISTANCE/ROTATE

```
DIMENSION XYZCA(4,10000),XYZOSN(3,10)
      CHARACTER*3 NAMECA(10000)
      ICA=1
11    READ(11,101,END=201)NAMECA(ICA),XYZCA(1,ICA),XYZCA(2,ICA)
      *,XYZCA(3,ICA)
101   FORMAT(13X,A3,14X,3F8.3)
      ICA=ICA+1
      GOTO 11
201   NCA=ICA-1
      XXCACN=0.
      XYCACN=0.
      XZCACN=0.
      DO 21 ICA=1,NCA
      XXCACN=XXCACN+XYZCA(1,ICA)
      XYCACN=XYCACN+XYZCA(2,ICA)
      XZCACN=XZCACN+XYZCA(3,ICA)
21    CONTINUE
      XXCACN=XXCACN/FLOAT(NCA)
      XYCACN=XYCACN/FLOAT(NCA)
      XZCACN=XZCACN/FLOAT(NCA)
      DO 22 ICA=1,NCA
      XYZCA(1,ICA)=XYZCA(1,ICA)-XXCACN
      XYZCA(2,ICA)=XYZCA(2,ICA)-XYCACN
      XYZCA(3,ICA)=XYZCA(3,ICA)-XZCACN
      XYZCA(4,ICA)=SQRT(XYZCA(1,ICA)**2+XYZCA(2,ICA)**2
      *+              XYZCA(3,ICA)**2)
22    CONTINUE
      WRITE(30,102)(NAMECA(ICA),(XYZCA(I,ICA),I=1,4),ICA=1,NCA)
102   FORMAT(5X,A3,3X,3F8.3,F10.3)
      STOP
      END
```

Appendix 4.3

Fortran source code for program Caldistance/rotate to search for match

between protein-bound calcium and hydroxyapatite calcium

This program sequentially centers the protein-bound calcium ions to each of the

two unique hydroxyapatite calcium and does rotation to find the best match.

```
DIMENSION XYZCA(3,10000),XYZOSN(3,10),OSNTMP(3,10),XCA(3,2)
      DEGTOR=ACOS(-1.)/180.
      ICA=1
C  READ IN HA SPHERE COORDINATE GENERATED BY PROGRAM LIKE SURFACE..ETC.
  11    READ(11,101,END=201)XYZCA(1,ICA),XYZCA(2,ICA)
      *,XYZCA(3,ICA)
  101   FORMAT(30X,3F8.3)
      ICA=ICA+1
      GOTO 11
  201   NCA=ICA-1
      XXCACN=0.
      XYCACN=0.
      XZCACN=0.
C  FIND CENTROID POSITION AND MOVE HA TO CENTER AT ORIGIN
      DO 21 ICA=1,NCA
      XXCACN=XXCACN+XYZCA(1,ICA)
      XYCACN=XYCACN+XYZCA(2,ICA)
      XZCACN=XZCACN+XYZCA(3,ICA)
  21    CONTINUE
      XXCACN=XXCACN/FLOAT(NCA)
      XYCACN=XYCACN/FLOAT(NCA)
      XZCACN=XZCACN/FLOAT(NCA)
      DO 22 ICA=1,NCA
      XYZCA(1,ICA)=XYZCA(1,ICA)-XXCACN
      XYZCA(2,ICA)=XYZCA(2,ICA)-XYCACN
      XYZCA(3,ICA)=XYZCA(3,ICA)-XZCACN
  22    CONTINUE
      IOSN=1
C      READ IN POSITION OF CA IN OSTEOCALCIN
C
  12    READ(12,101,END=202)XYZOSN(1,IOSN),XYZOSN(2,IOSN)
      *,XYZOSN(3,IOSN)
      IOSN=IOSN+1
      GOTO 12
```

```
  202   NOSN=IOSN-1
C       FILE 13 CONTAINS THE COORDINATES OF THE TWO
C       UNIQUE CALCIUM ATOMS. THIS IS GENERATED BY
C       PROGRAM CAL_CEN_MOL.F WHICH OUTPUT A FILE WITH
C       CENTERED HA WITH ATOM NAME,X,Y,Z, AND DISTANCE FROM
C       ORIGIN. TWO UNIQUE CA ATOMS CLOSEST TO ORIGIN ARE
C       SELECTED FOR GENERATING FILE 13
        READ(13,102)((XCA(J,I),J=1,3),I=1,2)
  102   FORMAT(30X,3F8.3)
        DO 301 IXCA=1,2
        DO 301 IOSN=1,NOSN
C       MOVE OSN WITH THIRD ATOM MOVED TO ORIGIN
C         THIS IS FOR KEEPING THE CENTROID CLOSE TO ORIGIN
C       ROTATE X, Y, Z IN SMALL INCREMENTS
C       AFTER EACH ROTATION, MOVE OSN WITH IOSN'S ATOM
C         TO OVERLAP WITH IXCA
C
C
C       MOVE OSN WITH IOSN'S ATOM MOVED TO ORIGIN
        DO 302 IO=1,NOSN
        OSNTMP(1,IO)=XYZOSN(1,IO)-XYZOSN(1,3)
        OSNTMP(2,IO)=XYZOSN(2,IO)-XYZOSN(2,3)
        OSNTMP(3,IO)=XYZOSN(3,IO)-XYZOSN(3,3)
C         WRITE(22,*)(OSNTMP(III,IO),III=1,3)
  302   CONTINUE
C       ROTATE X, Y, Z IN SMALL INCREMENTS
C       WE DO THIS BY ROTATING
C         1ST -- ROTATE ABOUT Z
C         2ND -- ROTATE ABOUT NEW X
C         3RD -- ROTATE ABOUT NEW Z
        DO 301 IROT1=0,355,5
        XROT1=FLOAT(IROT1)*DEGTOR
        DO 301 IROT2=0,355,5
        XROT2=FLOAT(IROT2)*DEGTOR
        DO 301 IROT3=0,355,5
        XROT3=FLOAT(IROT3)*DEGTOR
        XRMS=0.0
C   CALCULATE THE ROTATED POSITION OF
C     IOSN'S ATOM
        XXCEN=OSNTMP(1,IOSN)
        XYCEN=OSNTMP(2,IOSN)
        XZCEN=OSNTMP(3,IOSN)
        CALL ROTEUL(XXCEN,XYCEN,XZCEN,XROT1,XROT2,XROT3)
        DO 303 IO=1,NOSN
        XXX=OSNTMP(1,IO)
        XXY=OSNTMP(2,IO)
        XXZ=OSNTMP(3,IO)
        CALL ROTEUL(XXX,XXY,XXZ,XROT1,XROT2,XROT3)
C       AFTER EACH ROTATION, MOVE OSN WITH IOSN'S ATOM
C         TO OVERLAP WITH IXCA
C         WRITE(22,*)XXX,XXY,XXZ
        XXX=XXX+XCA(1,IXCA)-XXCEN
        XXY=XXY+XCA(2,IXCA)-XYCEN
        XXZ=XXZ+XCA(3,IXCA)-XZCEN
```

```
C        WRITE(6,*)XXX,XXY,XXZ
         XXDIST=10000000000000.
         DO 304 IXXCA=1,NCA
         XDIST= ((XXX-XYZCA(1,IXXCA))**2+
     *          (XXY-XYZCA(2,IXXCA))**2+
     *          (XXZ-XYZCA(3,IXXCA))**2)
         XXDIST=MIN(XDIST,XXDIST)
C        WRITE(22,*)XXDIST,XDIST,XRMS
  304 CONTINUE
         XRMS=XRMS+XXDIST
C        WRITE(22,*)XXDIST,XRMS

  303 CONTINUE
         XRMS=SQRT(XRMS)
         XCUT=2.0
         IF(XRMS.GE.XCUT) GOTO 301
         WRITE(21,401)IXCA,IOSN,IROT1,IROT2,IROT3,XRMS
  401 FORMAT(5I4,F12.3)
  301 CONTINUE
         STOP
         END


C
C    ROTATION MATRIX:   COUNTER CLOCKWISE POSITIVE
C          ROT X                 ROT Y               ROT Z
C    (1    0     0 )    (COS  0  SIN)    (COS  -SIN  0)
C    (0    COS -SIN)    ( 0   1   0 )    (SIN   COS  0)
C    (0    SIN +COS)    (-SIN 0  COS)    ( 0     0   1)
C  TO BE CONSISTENT WITH XPLOR, ANGLES ARE NEGATED
C  SO THE RESULT IS CLOCKWISE POSITIVE
         SUBROUTINE ROTX(X,Y,Z,THETA)
         COSTHE=COS(-THETA)
         SINTHE=SIN(-THETA)
         X2=X
         Y2=Y
         Z2=Z
         X=X2
         Y=Y2*COSTHE-Z2*SINTHE
         Z=Y2*SINTHE+Z2*COSTHE
         RETURN
         END
         SUBROUTINE ROTY(X,Y,Z,THETA)
         COSTHE=COS(-THETA)
         SINTHE=SIN(-THETA)
         X2=X
         Y2=Y
         Z2=Z
         Y=Y2
         Z=Z2*COSTHE-X2*SINTHE
         X=Z2*SINTHE+X2*COSTHE
         RETURN
         END
         SUBROUTINE ROTZ(X,Y,Z,THETA)
         COSTHE=COS(-THETA)
```

```
SINTHE=SIN(-THETA)
X2=X
Y2=Y
Z2=Z
Z=Z2
X=X2*COSTHE-Y2*SINTHE
Y=X2*SINTHE+Y2*COSTHE
RETURN
END
SUBROUTINE ROTEUL(X,Y,Z,T1,T2,T3)
CALL ROTZ(X,Y,Z,T1)
CALL ROTX(X,Y,Z,T2)
CALL ROTZ(X,Y,Z,T3)
RETURN
END
```

Appendix 4.4

Fortran source code for program OUT_ALL to sort and output only the

solutions within a certain cutoff

This program allows the user to select the RMSD cutoff and output the rotations

lower than the cutoff.

```
WRITE(6,101)
 101   FORMAT(' ENTER XDIST CUTOFF')
       READ(5,*)XCUT
 1     READ(21,*,END=201)I1,I2,I3,I4,I5,XDIST
       IF(XDIST.GT.XCUT)GOTO 1
       WRITE(23,301)I1,I2,I3,I4,I5,XDIST
 301   FORMAT(5I4,F8.3)
       GOTO 1
 201   STOP
       END
```

Appendix 4.5

Fortran source code for program ROT2_FINAL to fine search for match

between protein-bound calcium and hydroxyapatite calcium

```
C   PROGRAM ROT12_final
C   THIS PROGRAM IS MODIFIED FORM OF CALDISTANCE
C    IT READ OUTPUT FILE FROM CALDISTANCE
C      FORT.21 AFTER SORTED AND RENAME TO FORT.14
C
      DIMENSION XYZCA(3,10000),XYZOSN(3,10),OSNTMP(3,10),XCA(3,2)

      DIMENSION IXCAXX(200),IOSNXX(200),IR1XX(200),IR2XX(200),
     *          IR3XX(200),XDSTXX(200),IRR1(200),IRR2(200),
     *   IRR3(200),ITT1(200),ITT2(200),ITT3(200)
      CHARACTER*30 LINE(10),LINN
      DEGTOR=ACOS(-1.)/180.
      XONE=1.0
      NOUT=31
      ICA=1
C   READ IN HA SPHERE COORDINATE GENERATED BY PROGRAM LIKE SURFACE..ETC.
   11     READ(11,101,END=201)LINN,XYZCA(1,ICA),XYZCA(2,ICA)
     *,XYZCA(3,ICA),XJUNK,XJUNK
  101     FORMAT(A30,3F8.3,2F6.2)
          ICA=ICA+1
          GOTO 11
  201     NCA=ICA-1
          XXCACN=0.
          XYCACN=0.
          XZCACN=0.
C   FIND CENTROID POSITION AND MOVE HA TO CENTER AT ORIGIN
          DO 21 ICA=1,NCA
          XXCACN=XXCACN+XYZCA(1,ICA)
          XYCACN=XYCACN+XYZCA(2,ICA)
          XZCACN=XZCACN+XYZCA(3,ICA)
   21     CONTINUE
          XXCACN=XXCACN/FLOAT(NCA)
          XYCACN=XYCACN/FLOAT(NCA)
          XZCACN=XZCACN/FLOAT(NCA)
          DO 22 ICA=1,NCA
          XYZCA(1,ICA)=XYZCA(1,ICA)-XXCACN
          XYZCA(2,ICA)=XYZCA(2,ICA)-XYCACN
          XYZCA(3,ICA)=XYZCA(3,ICA)-XZCACN
   22     CONTINUE
          IOSN=1
```

```
C        READ IN POSITION OF CA IN OSTEOCALCIN
C
  12     READ(12,101,END=202)LINE(IOSN),XYZOSN(1,IOSN)
        *,XYZOSN(2,IOSN)
        *,XYZOSN(3,IOSN)
         IOSN=IOSN+1
         GOTO 12
 202     NOSN=IOSN-1
C     FILE 13 CONTAINS THE COORDINATES OF THE TWO
C     UNIQUE CALCIUM ATOMS. THIS IS GENERATED BY
C     PROGRAM CAL_CEN_MOL.F WHICH OUTPUT A FILE WITH
C     CENTERED HA WITH ATOM NAME,X,Y,Z, AND DISTANCE FROM
C     ORIGIN. TWO UNIQUE CA ATOMS CLOSEST TO ORIGIN ARE
C     SELECTED FOR GENERATING FILE 13
         READ(13,102)((XCA(J,I),J=1,3),I=1,2)
 102     FORMAT(30X,3F8.3)
         ISOL=1
  1      READ(15,*,END=104)IXCAXX(ISOL),IOSNXX(ISOL),
        *IR1XX(ISOL),IR2XX(ISOL),IR3XX(ISOL),XDSTXX(ISOL)
        *,IRR1(ISOL),IRR2(ISOL),IRR3(ISOL)
        *,ITT1(ISOL),ITT2(ISOL),ITT3(ISOL)
         ISOL=ISOL+1
         GOTO 1
 104     NSOL=ISOL-1
C        DO 301 ISOL=1,NSOL
         DO 301 ISOL=1,NSOL

C        DO 301 IXCA=1,2
C        DO 301 IOSN=1,NOSN
C     RESTORE IXCA AND IOSN FROM STORED INPUT VALUE
         IXCA=IXCAXX(ISOL)
         IOSN=IOSNXX(ISOL)

C        MOVE OSN WITH THIRD ATOM MOVED TO ORIGIN
C           THIS IS FOR KEEPING THE CENTROID CLOSE TO ORIGIN
C        ROTATE X, Y, Z IN SMALL INCREMENTS
C        AFTER EACH ROTATION, MOVE OSN WITH IOSN'S ATOM
C          TO OVERLAP WITH IXCA
C
C
C        MOVE OSN WITH IOSN'S ATOM MOVED TO ORIGIN
         DO 302 IO=1,NOSN
         OSNTMP(1,IO)=XYZOSN(1,IO)-XYZOSN(1,3)
         OSNTMP(2,IO)=XYZOSN(2,IO)-XYZOSN(2,3)
         OSNTMP(3,IO)=XYZOSN(3,IO)-XYZOSN(3,3)
C         WRITE(22,*)(OSNTMP(III,IO),III=1,3)
 302     CONTINUE
C        ROTATE X, Y, Z IN SMALL INCREMENTS
C        WE DO THIS BY ROTATING
C         1ST -- ROTATE ABOUT Z
C         2ND -- ROTATE ABOUT NEW X
C         3RD -- ROTATE ABOUT NEW Z
C         DO 301 IROT1=0,355,5
         IROT1=IR1XX(ISOL)
```

```
          XROT1=FLOAT(IROT1)*DEGTOR
C         DO 301 IROT2=0,355,5
          IROT2=IR2XX(ISOL)
          XROT2=FLOAT(IROT2)*DEGTOR
C         DO 301 IROT3=0,355,5
          IROT3=IR3XX(ISOL)
          XROT3=FLOAT(IROT3)*DEGTOR

C CARRY OUT A FINE ROTATION SEARCH AROUND XROT1,XROT2,XROT3)
C         DO 401 IR1=-15,15,5
          IR1=IRR1(ISOL)
C         IR1=0
          XXROT1=XROT1+DEGTOR*0.1*FLOAT(IR1)
C         DO 401 IR2=-15,15,5
          IR2=IRR2(ISOL)
C         IR2=0
          XXROT2=XROT2+DEGTOR*0.1*FLOAT(IR2)
C         DO 401 IR3=-15,15,5
          IR3=IRR3(ISOL)
C         IR3=0
          XXROT3=XROT3+DEGTOR*0.1*FLOAT(IR3)

C
C    CARRY OUT A FINE TRANSLATION SEARCH
C    FROM -1.5 ANGSTROM to +1.5 ANGSTROM
C
C         DO 401 ITRAN1=-15,15,5
          ITRAN1=ITT1(ISOL)
C         ITRAN1=0
          XT1=0.1*FLOAT(ITRAN1)
C         DO 401 ITRAN2=-15,15,5
          ITRAN2=ITT2(ISOL)
C         ITRAN2=0
          XT2=0.1*FLOAT(ITRAN2)
C         DO 401 ITRAN3=-15,15,5
          ITRAN3=ITT3(ISOL)
C         ITRAN3=0
          XT3=0.1*FLOAT(ITRAN3)


          XRMS=0.0

C   CALCULATE THE ROTATED POSITION OF
C     IOSN'S ATOM
          XXCEN=OSNTMP(1,IOSN)
          XYCEN=OSNTMP(2,IOSN)
          XZCEN=OSNTMP(3,IOSN)
          CALL ROTEUL(XXCEN,XYCEN,XZCEN,XXROT1,XXROT2,
         *XXROT3)




          DO 303 IO=1,NOSN
```

```
      XXX=OSNTMP(1,IO)
      XXY=OSNTMP(2,IO)
      XXZ=OSNTMP(3,IO)
      CALL ROTEUL(XXX,XXY,XXZ,XXROT1,XXROT2,
     *XXROT3)
C     AFTER EACH ROTATION, MOVE OSN WITH IOSN'S ATOM
C       TO OVERLAP WITH IXCA
C      WRITE(22,*)XXX,XXY,XXZ
      XXX=XXX+XCA(1,IXCA)-XXCEN+XT1
      XXY=XXY+XCA(2,IXCA)-XYCEN+XT2
      XXZ=XXZ+XCA(3,IXCA)-XZCEN+XT3
C      WRITE(6,*)XXX,XXY,XXZ
      XXDIST=10000000000000.
      DO 304 IXXCA=1,NCA
      XDIST= ((XXX-XYZCA(1,IXXCA))**2+
     *        (XXY-XYZCA(2,IXXCA))**2+
     *        (XXZ-XYZCA(3,IXXCA))**2)
      XXDIST=MIN(XDIST,XXDIST)
C      WRITE(22,*)XXDIST,XDIST,XRMS
  304 CONTINUE
      XRMS=XRMS+XXDIST
C      WRITE(22,*)XXDIST,XRMS
      WRITE(NOUT,101)LINE(IO),XXX,XXY,XXZ,XONE,
     *XDSTXX(ISOL)

  303 CONTINUE
      XRMS=SQRT(XRMS)
      WRITE(24,305)IXCA,IOSN,IROT1,IROT2,IROT3,XRMS
     *,IR1,IR2,IR3,ITRAN1,ITRAN2,ITRAN3
  305 FORMAT(5I4,F8.3,6I4)
      WRITE(25,306)XRMS
  306  FORMAT(F10.3)
C 401  CONTINUE
      NOUT=NOUT+1
  301 CONTINUE
      STOP
      END

C
C   ROTATION MATRIX:  COUNTER CLOCKWISE POSITIVE
C       ROT X              ROT Y            ROT Z
C    (1   0    0 )   (COS  0  SIN)   (COS  -SIN  0)
C    (0   COS -SIN)  ( 0   1   0 )   (SIN   COS  0)
C    (0   SIN +COS)  (-SIN 0  COS)   ( 0     0   1)
C  TO BE CONSISTENT WITH XPLOR, ANGLES ARE NEGATED
C  SO THE RESULT IS CLOCKWISE POSITIVE
      SUBROUTINE ROTX(X,Y,Z,THETA)
      COSTHE=COS(-THETA)
      SINTHE=SIN(-THETA)
      X2=X
      Y2=Y
      Z2=Z
      X=X2
      Y=Y2*COSTHE-Z2*SINTHE
```

```
Z=Y2*SINTHE+Z2*COSTHE
RETURN
END
SUBROUTINE ROTY(X,Y,Z,THETA)
COSTHE=COS(-THETA)
SINTHE=SIN(-THETA)
X2=X
Y2=Y
Z2=Z
Y=Y2
Z=Z2*COSTHE-X2*SINTHE
X=Z2*SINTHE+X2*COSTHE
RETURN
END
SUBROUTINE ROTZ(X,Y,Z,THETA)
COSTHE=COS(-THETA)
SINTHE=SIN(-THETA)
X2=X
Y2=Y
Z2=Z
Z=Z2
X=X2*COSTHE-Y2*SINTHE
Y=X2*SINTHE+Y2*COSTHE
RETURN
END
SUBROUTINE ROTEUL(X,Y,Z,T1,T2,T3)
CALL ROTZ(X,Y,Z,T1)
CALL ROTX(X,Y,Z,T2)
CALL ROTZ(X,Y,Z,T3)
RETURN
END
```