AN ABSTRACT OF THE THESIS OF

K. Shawn Watts for the degree of Master of Science in Biochemistry and Biophysics presented on July 31, 2000. Title: Steps Toward Structure-Assisted Drug Design.

Redacted for Privacy

Abstract approved: _____

Pui Shing Ho

The three dimensional structure of both a ligand and its cognate receptor are required for the success of structure-assisted drug design. This thesis reports the crystal structure of hectochlorin, a small, bioactive molecule, and the steps toward determining the crystal structure of an RNA molecule that is an attractive target for drug design.

The absolute structure of hectochlorin, a cytotoxic, secondary metabolite isolated from *Lyngbya majuscula*, is reported herein. Specifically, the absolute configuration of hectochlorin, as determined by x-ray crystallography, is reported as 6S, 7S, 10S, 31S. Marine natural products are interesting as a source of novel chemical compounds that are potentially valuable as therapeutic agents, or have industrial applications. The absolute structure provides a model that serves as a starting point for rational drug design synthesis.

In a second study, results are reported from attempts to crystallize a biologically important RNA structure, the *trans*-acting response element, (TAR), for the determination of its structure by x-ray diffraction, and ultimately, providing an initial model for structure-assisted drug design targeted against HIV. Crystals, of biologically relevant TAR sequences, greater that 0.1 x 0.1 x 0.1

mm^3 in size, both in the presence and absence of a cognate ligand analogue, have been obtained. These crystals have been shown to be of poor diffraction quality, but the initial crystallization conditions provide a starting point for optimization that may yield higher quality crystals.

Steps Toward Structure-Assisted Drug Design

by

K. Shawn Watts

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented July 31, 2000
Commencement June 2001

Master of Science thesis of K. Shawn Watts presented on July 31, 2000

APPROVED:

Redacted for Privacy

Major Professor, representing Biochemistry and Biophysics

Redacted for Privacy

Chair of Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

K. Shawn Watts, Author

ACKNOWLEDGMENT

I thank my major professor P. Shing Ho and committee members, Phil Proteau, Mike Schimerlik, and Mike Penner, for their help and guidance in writing this thesis. I especially thank Brian Márquez for his, encouragement, friendship, and providing the hectochlorin crystals.

I also express heartfelt gratitude to lab colleagues, Brandt Eichman and Jeff Vargason, for their careful and constant tutelage of x-ray crystallography and science in general. I am most grateful to the many members of the Department of Biochemistry and Biophysics for allowing me to learn amongst them. Finally, I thank Alex Yokochi for his instruction of point counter data collection and SHELX refinement

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Steps Toward Structure-Assisted Drug Design

Chapter 1 Structure-assisted Drug Design

INTRODUCTION

As the understanding of biomolecular structure advances, the hope of predicting accurate structure-activity relationships becomes more tantalizing. The paradigm of structural biology dictates that the structure of a molecule determines its function. At this time, however, the ability to successfully predict a molecule's function solely from its structure is feeble. Nowhere else will strengthening this ability to predict activity from structure have greater impact than on structure-based drug design. This thesis deals with two specific examples of steps towards structure-based drug design. Chapter 2 presents the determination of the absolute configuration of hectochlorin, a natural product, having potential therapeutic applications, for the purpose of simplifying its synthesis. Chapter 3 describes the results of attempts to crystallize, and ultimately determine the structure of an RNA molecule critical to the human immunodeficiency virus life cycle, which is a potential drug target.

Only in recent history, have bridges begun to form between the awareness of an active compound in a natural product, and the structural rationale for its therapeutic effects. Many plant-based remedies have long been used before the underlying science for its action was understood. Salicin

from the willow tree, used as a mild analgesic, is just one example. The compound was isolated in 1829, but not until 1899, was the acetate ester of salicylic acid introduced by Bayer, as aspirin (1). Only much later was aspirin's action as a cyclooxygenase inhibitor discovered.

Historically, a newly discovered active compound was optimized through rational drug development. Classic rational drug development is an iterative process that develops a lead compound into a drug candidate. In this process, a compound with some desired activity is selected and a library of derivatives is created. By screening a batch of derivatives for improved activity, a new best candidate is selected and derivatized again. The aim of this process is to enhance desired activities while reducing undesirable properties. This process is tried and true, albeit slow. Typically, thousands of compounds are created and rejected along the path that leads to approval for a single drug; this process frequently takes over a decade.

Structure based drug design offers an alternative to the tedious screening of large libraries of compounds, looking for the proverbial needle in the haystack. Structure based drug design is the ambitious concept that the structure of a target receptor contains all the information required to design an ideal ligand for that receptor. This approach can potentially be applied to a broad spectrum of diseases. A crystal structure offers detailed positions of atoms that make up the receptor and the surrounding solvent. Detailed knowledge of the receptor's structure can guide the design of complementary ligands that optimize favorable contacts and eliminate sterically unfavorable

geometries within a bound complex. Structure based design of ligands will theoretically result in tighter ligand binding, as evident from lower dissociation constants. A tight binding ligand translates to a lower dose for an equivalent desired effect; lower dose would potentially reduce concentration dependent side effects.

Using structure-assisted drug design, it is theoretically possible to make modifications to a ligand that will not only augment binding activity, but also enhance other properties of the compound. A pro-drug is a modification to a therapeutic agent's structure that, in addition to masking activity until unveiled by metabolism, can improve other properties of the drug. There are many qualities of a ligand that are desirable to control, such as a its solubility, cellular uptake, tissue selectivity and metabolism. It is possible, for example, to exchange substituents that are inconsequential to binding activity for functional groups that are platforms for pro-drug development.

A true success for structure-assisted drug design resulted from work on human immunodeficiency virus protease (HIV PR). Structure-assisted drug design lead to development of novel compounds such as DMP-450. DMP-450 is C_2 symmetric and is designed to complement the C_2 symmetry of the HIV protease receptor. Additionally, a keto oxygen in DMP-450 is designed to displace a structural water found in HIV PR. Both concepts, the symmetry and the incorporation of the keto oxygen, were derived directly from analysis of the receptor's three dimensional structure. This compound shows

remarkable specificity to HIV PR, is a potent inhibitor, and is resistant to protease degradation (2, 3).

Perhaps more importantly, analysis of the HIV PR structure and libraries of small molecule structures facilitated the discovery of new lead compounds through computational screening. Using programs to sift through libraries of small molecules , and to study their ability to dock into the invaginations of the receptor site of HIV PR, has produced a number of new lead compounds. These new compounds are unrelated to first generation peptide mimics. This provides novel frameworks upon which new families of inhibitors can be designed. In many cases, new lead compounds were determined with mere hours of computation. This ability to extract novel lead compounds that are selective for a specific binding site is a major achievement in structural biology (4, 5).

One example of where future structure-assisted drug design may facilitate the development of a natural product into a drug is psoralen, a DNA photocrosslinking agent. The plant source of this compound, *Psoralea corylifolia*, is described in antique remedies for vitiligo as early as 1500 BC. Not until the 1940s did scientists isolate the active compounds, now known as psoralens. The base structure of psoralen was later determined to be tricyclic and planar (6, 7).

Psoralen accumulates in unhealthy skin cells and tissue with increased vascular flow. In the 1970s psoralen was shown to non-specifically intercalate into DNA duplexes. When irradiated with light of 320-400 nm, the psoralen

intercalated between T-A steps becomes covalently crosslinked to the thymine residues (6, 7).

Current research is being directed to develop therapies for cancer and skin disease utilizing psoralen's photocrosslinking properties. An immediate obstacle for this strategy is that the wavelengths that promote crosslinking do not penetrate tissues effectively (7). However, longer wavelength radiation has better tissue penetration properties. To increase the effectiveness of the therapy, the parent structure of psoralen may be modified with substituents that extend conjugation and hyperconjugation. This approach may extend the wavelength at which the pi system is excited.

Recently, the atomic resolution structure of a psoralen compound cross-linked to DNA has been solved (Eichman *et al.*, manuscript submitted). This structure may hold keys to overcoming the deficiency of psoralen photocrosslinking therapy. Substituent modifications to psoralen that extend the wavelengths required for cross-linking, yet don't interfere with critical elements of DNA binding, may result from analysis of this structure.

As biology is a chiral realm, any drug design strategy must be concerned with stereochemical details. A complete stereochemical description of the ligand is integral to predicting interactions with the receptor site. Potentially, the mere presence of required functional groups on a ligand is the sole requirement for activity. However, it is more likely that the configuration of functional groups is equally important to forming a cognate to the receptor's

structure. Therefore, enantiomeric relationships are an important feature of a ligand's structure activity relationships.

Enantiomers of drugs can have unanticipated effects. For example, (R)-11-hydroxy-10-methylaporphine is a selective agonist of 5-hydroxytryptamine receptor 1A (5-HT_{1A}). In contrast, (S)-11-hydroxy-10-methylaporphine is an antagonist for the same 5-HT_{1A} receptor. Furthermore, one enantiomer can be primarily responsible for undesirable side effects in a racemic mixture. Thalidomide is an often cited example of one enantiomer being primarily responsible for unwanted side effects. Animal tests, utilizing SWS mice, show both enantiomers of thalidomide possess hypnotic effects but the S enantiomorph is culpable for teratogenic side effects (8).

One of the most powerful tools in structure-assisted drug design is x-ray crystallography. X-ray crystallography is an excellent tool for describing the structure of a receptor's binding site and a ligand's structure in great detail. This technique provides a three dimensional electron density map of molecules in a single crystal. Greater map resolution translates into higher confidence in the exact position of the atoms in the molecule.

PROBLEM DEFINITION

The structure of ligand and receptor must be accurately defined for structure-assisted drug design to be applicable. X-ray crystallography is a

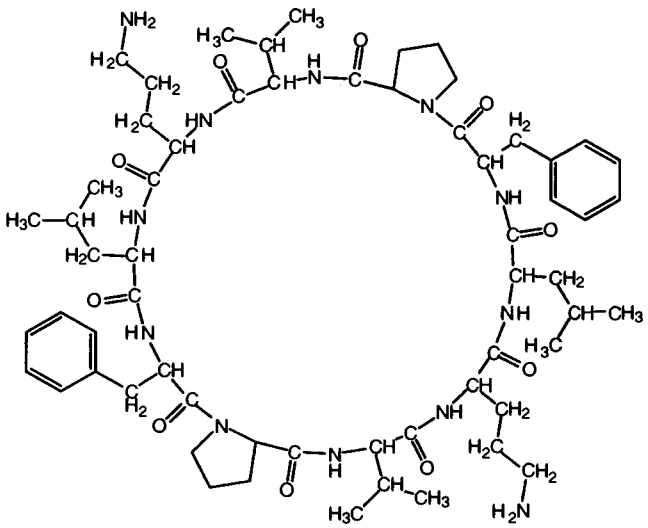
powerful technique that provides this information. The following chapters describe two examples of crystallography in drug design. Chapter 2 details the absolute structure, determined by single-crystal x-ray diffraction, of hectochlorin, a ligand with multiple stereocenters, and discusses possible biogenesis pathways for the compound. Chapter 3 focuses on attempts to crystallize a biologically important RNA structure, the *trans*-acting response element, (TAR), for the determination of its structure by x-ray diffraction, and, thus, provide a starting point for structure based drug design to target the AIDS virus.

Chapter 2 The Absolute Structure of Hectochlorin

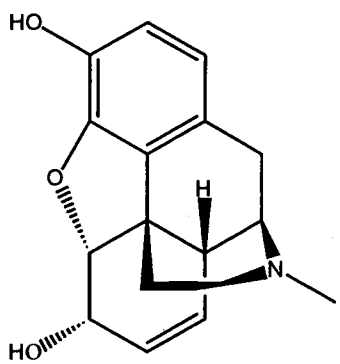
INTRODUCTION

Hectochlorin, **1**, is a cytotoxic secondary metabolite isolated from *Lyngbya majuscula* that contains four stereogenic carbons. A secondary metabolite is a molecule that may not be required for an organism to grow and reproduce, but potentially confers some advantage for survival (9). In the case of secondary metabolite-rich organisms such as terrestrial plants, bacteria, and marine algae, the secondary metabolites may act as a deterrent to predators or competitors. Interestingly, these metabolites can be useful in the treatment of human disease or may have industrial applications. Therefore, they are fervently investigated. Figure 2.1 shows the structure of four secondary metabolites with medicinal or industrial applications, as classified by their biosynthetic pathway.

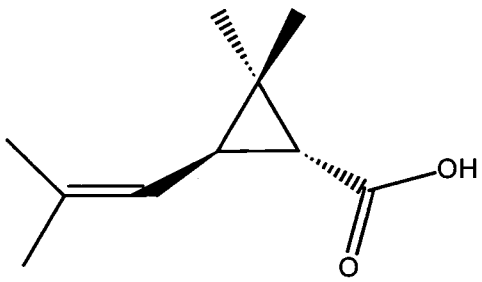
Marine organisms are a rich source of novel compounds with potential value. Curacin A is an example of a marine natural product with cancer fighting properties. The molecule is isolated from a *Lyngbya majuscula* harvested from the coastal waters of Curaçao. This compound is a potent inhibitor of microtubule formation, with an IC_{50} value in the nanomolar range. Furthermore, the compound demonstrates selective activity against breast, colon, and renal cancer cell lines (10).



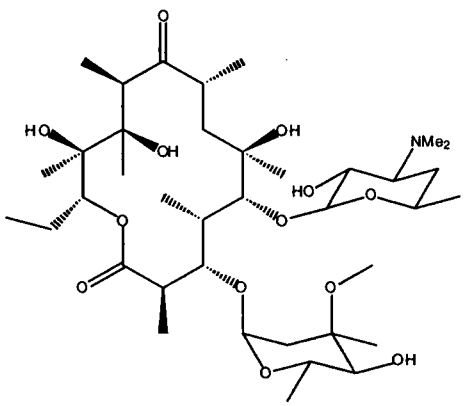
Gramicidin S



Morphine



Chrysanthemic acid



Erythromycin A

Figure 2.1 Secondary metabolites classed by their biosynthetic pathways. This figure shows the structures of compounds representative of several classes of biosynthesis: alkaloids, terpenes, polyketides, and non-ribosomal polypeptides. Each has medicinal or industrial applications. Chrysanthemic acid is a terpenoid moiety of an insecticide, erythromycin A is an antibiotic polyketide, Gramicidin S is a non-ribosomal polypeptide antibiotic

As further examples of marine natural products, consider *cis, cis-*, and *trans, trans*-ceratospongamide. These compounds are isolated from *Ceratodictyon spongiosum/sigmadocia symbiotica*, a red alga and sponge symbiont. *Trans, trans*-ceratospongamide is an anti-inflammatory response agent. Secreted phospholipase ($sPLA_2$) catalyzes the release of arachidonic acid from membrane glycophospholipids. Arachidonic acid is modified into eicosanoids that elicit an inflammation response. *Trans, trans*-ceratospongamide is shown to effectively inhibit the expression of $sPLA_2$ with an ED_{50} of 32 nM. Thus, it limits the amount of arachidonic acid available for eicosanoid synthesis and reduces the inflammation response. The *cis, cis*-ceratospongamide isomer, however, is inactive in this assay. This demonstrates that subtle aspects of structure are critical to the activity of a compound (11).

Hectochlorin, isolated from *Lyngbya majuscula*, showed cytotoxicity in brine shrimp, and snail assays. Additionally, a fungal assay with *Candida albicans* showed a 10.5 mm zone of inhibition at 10 μg/mL. This activity encourages further analysis and structure determination. Prior investigation of **1** using NMR and MS revealed the molecular formula and the relative structure containing four stereocenters. The four stereogenic carbons lead to a total of sixteen possible stereoisomers. To completely characterize **1**, the absolute stereochemistry has to be unambiguously defined.

In this study, x-ray crystallography, incorporating anomalous scattering data, was used to resolve the enantiomorph ambiguity. **1** contains two

chlorine and two sulfur atoms which contribute sufficient anomalous scattering, relative to the remaining scattering mass, to determine the absolute configuration of the molecule. The compound readily provided single crystals for absolute structure determination using x-ray crystallography. Each stereocenter is clearly defined from the single crystal study, thus completing characterization of the compound's absolute stereochemical structure.

Given that **1** readily generates large diffraction quality crystals, x-ray diffraction is an ideal tool for confirming relative stereochemistry. The heavy atoms, sulfur and chlorine, provide the important advantage of anomalous scattering for absolute stereochemistry determination. Because the number of atoms in **1** is small, direct methods were used to solve the phase problem in this experiment. The purpose of this chapter is to describe this work and the resulting absolute structure of hectochlorin.

METHODS AND MATERIALS

Hectochlorin formed colorless crystals from equal parts of water and methanol. A 0.3 x 0.3 x 0.3 mm^3 crystal was mounted in a glass capillary tube. Graphite monochromated Cu K_α radiation from a Siemens P4 sealed tube source was used to record 5813 reflections representing data out to 0.85 Å resolution. XSCANS (Siemens) employed 97 reflections stronger than 25 σ to

index the unit cell as $P2_12_12_1$. The space group was confirmed by examination of systematic reflection absences along Miller indices.

In an x-ray diffraction experiment the intensities of x-rays diffracted from a single crystal are carefully measured. The dimensions and space group symmetry of the crystal's unit cell define the diffraction pattern, a constellation of so-called "reflections". The individual reflections in the pattern contain information about the contents of the unit cell. Unfortunately, the measured reflection intensities alone do not afford the relative positions of the atoms in the unit cell.

The complete structure factor, F(S), for each reflection contains information on the scattering mass, f, and relative positions of each atom in a unit cell (Eq 1). In Equation 1, $S \bullet r_j$ is the phase angle contribution to a single reflection from a single atom, with r_j as the vector describing the atomic position of the j^{th} atom.

$$F(h, k, l) = F(S) = \sum_{j=1}^{N} f_j \exp 2\pi i(S \bullet r_j) \quad \text{Eq (1)}$$

A three dimensional map of the electron density is generated by a summation of the structure factors, as in Equation (2).

$$\rho(r) = (1/V) \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(h, k, l) \exp -2\pi i(S \bullet r) \quad \text{Eq (2)}$$

To generate a correct and interpretable map of the unit cell, the phase angle of each reflection must also be defined. The intensity, but not the phase, of each

reflection is measured in a single experiment. Therefore, the phase must be determined separately. This is known as the phase problem in X-ray crystallography (12).

Direct methods are a strictly mathematical approach to solving the phase problem for a structure with a small number of atoms. Direct methods result in an initial description of the electron density map, and hence, approximate atomic positions. Three phases of normalized structure factors from centric reflections, reflections that may adopt only two possible phases, are arbitrarily defined to select an origin and coordinate system. Sets of related normalized structure factors are chosen to estimate initial phases using the triplet relationship, Equation (3) (13).

$$0 \cong \Phi(h_1, k_1, l_1) + \Phi(h_2, k_2, l_2) + \Phi(-(h_1, k_1, l_1,) + (h_2, k_2, l_2,)) \quad \text{Eq (3)}$$

Starting from defined phases used to fix the origin and expanding the number of estimated phases from the triplet relationship, the entire set of reflections is phased sufficiently well to provide a first estimate density map. SHELXS is a direct methods program for solving the phase problem for small molecules (14). The output from SHELXS is a list of electron density peak positions derived from these triplet relationships. A model is generated by assigning the peaks in the initial map to atoms in the structure using chemical intuition and all available information. For example, the highest electron density peaks are designated as atoms with the most electrons in the chemical formula.

The initial model from direct methods is subsequently improved using a least-squares matrix refinement. The square of the disagreement between $F(h, k, l)_{obs}$ and $F(h, k, l)_{calc}$ is minimized by altering $F(h, k, l)_{calc}$. $F(h, k, l)_{calc}$ is composed of a number of modeled parameters. For example, the x, y, and z position of all the atoms and the individual B-factors are constituent variables of $F(h, k, l)_{calc}$. The individual parameters are shifted slightly and the value of the disagreement between $F(h, k, l)_{obs}$ and $F(h, k, l)_{calc}$ is evaluated. When the value of each parameter in $F(h, k, l)_{calc}$ is slightly shifted yet the disagreement function remains at a minimum, parameters are presumed refined. SHELXL used to refine the atomic positions and B-factors in a model (14).

Even the correct assignment of the phase magnitudes in a refined model does not provide absolute stereochemistry by itself. The sign of the phase angles determines the "handedness" of the electron density map. The relative stereochemistry is all that could be gleaned from an x-ray diffraction experiment without measurable anomalous scattering as described below (15).

With the correct magnitude of the phase angle applied to the structure factor of each reflection, the resulting map shows an accurate distribution of electron density. However, a change in sign to each phase angle will generate an equally accurate electron density map of the mirror image of the contents of the unit cell. Determining the "handedness" of the map is an important

consideration. For many systems, where stereochemistry is selected by nature and known, the map "handedness" must agree with the natural stereoisomer. Thus, in order to accept the signs of the phases, an electron density map of a ribosomal product protein must show right-handed α-helices with L-amino acids, and a map of a biosynthesized DNA must show D-ribose sugars. In the case of a newly discovered natural product, however, the stereochemistry may not be prescribed. Many natural products have D-amino acids and incorporate a variety of other chiral moieties that require definition. Anomalous scattering data allows direct determination of the correct configuration of a natural product by X-ray crystallography.

The lengths and angles of the unit cell, with the right hand rule, define the assignment of the Miller indices to all reflections. In a theoretical array of scattering atoms, Friedel's law states that the phase angle for a reflection with Miller indices (h, k, l) will have the same magnitude but opposite sign as the reflection with Miller indices (-h, -k, -l). The resultant beam scattering from a theoretical free electron, is 180 degrees out of phase from the incident beam (12, 13).

In contrast, the electrons in real systems are associated with atoms in quantized levels. Interactions between incident x-rays and electrons in real systems, such as absorption or fluorescence, can result in small changes in the magnitude and phase of the resultant structure factors. A break down of Friedel's law results from these small changes. The effect of anomalous scattering on the magnitude and phase of pair of structure factors is shown in

Figure 2.2. Anomalous scattering allows the determination of the correct enantiomorph by comparing the structure factors observed for a Friedel pair with the calculated structure factors for a particular enantiomer model (12, 13).

The program SHELXL can clearly distinguish between enantiomers using anomalous scattering data. The enantiomer of the model is generated, and the program queries how much of each enantiomer must be present to account for observed data. If this value converges near 1.0 or 0.0 then the crystal is enantiomerically pure. A returned value of 0.0 indicates data best fits the original enantiomer model. A value of 1.0 suggests the mirror image is best represented by the data. In this manner the absolute structure is clearly determined (14).

The structure of hectochlorin was solved and refined with SHELXS and SHELXL, respectively (14). A single hectochlorin molecule and one solvent molecule constituted the asymmetric unit (ASU) of the crystal.

RESULTS AND DISCUSSION

The purpose of this study was to reveal the absolute structure of **1**. 6S, 7S, 10S, 31S defines the stereochemistry. The data collection and refinement statistics for hectochlorin are given in Table 2.1, the absolute structure is shown in Figure 2.3, and an ORTEP representation of the ASU is shown in Figure 2.4. The complete description of a molecular structure must include

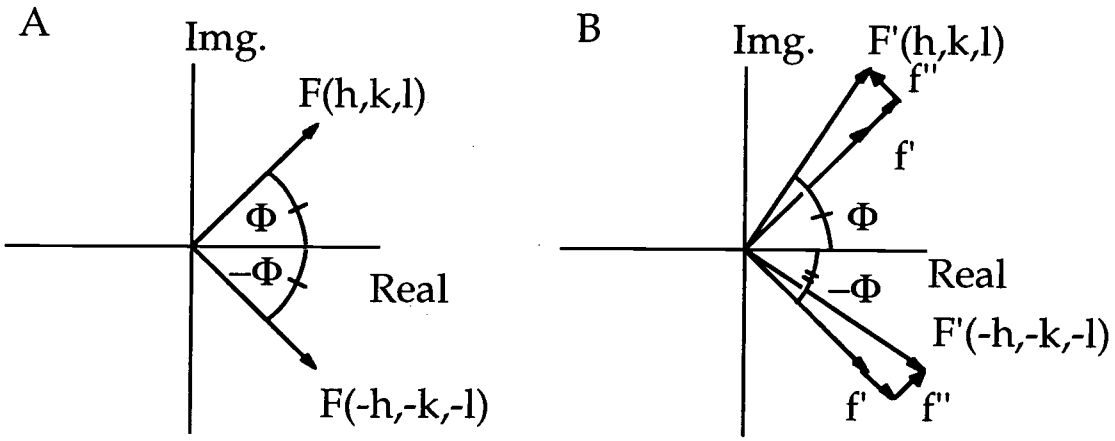


Figure 2.2 The effect of anomalous scattering on Friedel mates. Panel A is an Argand diagram of a structure factor F(h, k, l) and its Friedel mate F(-h,-k,-l) without anomalous scattering. The vertical axis is the imaginary component and the horizontal axis is the real component of the structure factor. The angle from the positive real axis to the structure factor, F(h,k,l), is the phase angle, Φ. From A, it is clear that Friedel's law is obeyed, and the magnitude of the structure factor intensities and phases are identical; only the sign of the phase is different. In panel B, the same structure factors are presented with anomalous scattering, where Friedel's law no longer holds true. The anomalous intensity magnitude is represented by the vector f'. The anomalous change in phase angle is represented by the vector f''. The resultant structure factors F'(h, k, l) and F'(-h,-k,-l) are therefore distinguishable by intensity magnitude and phase angle (12, 13).

Table 2.1 Data collection and refinement statistics for hectochlorin. Space group and unit cell dimensions are reported along with refinement statistics. Definitions of refinement values are included below table.

Space Group	$P2_12_12_1$
Unit Cell	
a, Å	12.266
b, Å	12.684
c, Å	21.415
$\alpha = \beta = \gamma$, degrees	90
Refinement parameters*	
R	6.66 % (5.59 %)
wR^2	15.33 % (14.33 %)
GooF for 407 parameters	1.043

*The R and wR^2 values in parenthesis are for reflections with $F^2 > 4\sigma$, the refinement parameters are defined below.

$R = \Sigma \mid \mid (F_{obs}) \mid - \mid (F_{calc}) \mid \mid / \Sigma \mid F_{obs} \mid$

$wR^2 = \{ \Sigma (w(F_{obs}^2 - F_{calc}^2)^2) / \Sigma (w (F_{obs}^2)^2) \}^{1/2}$

GooF (Goodness of Fit) = $S = \{ \Sigma (w(F_{obs}^2 - F_{calc}^2)^2) / (n - p) \}^{1/2}$

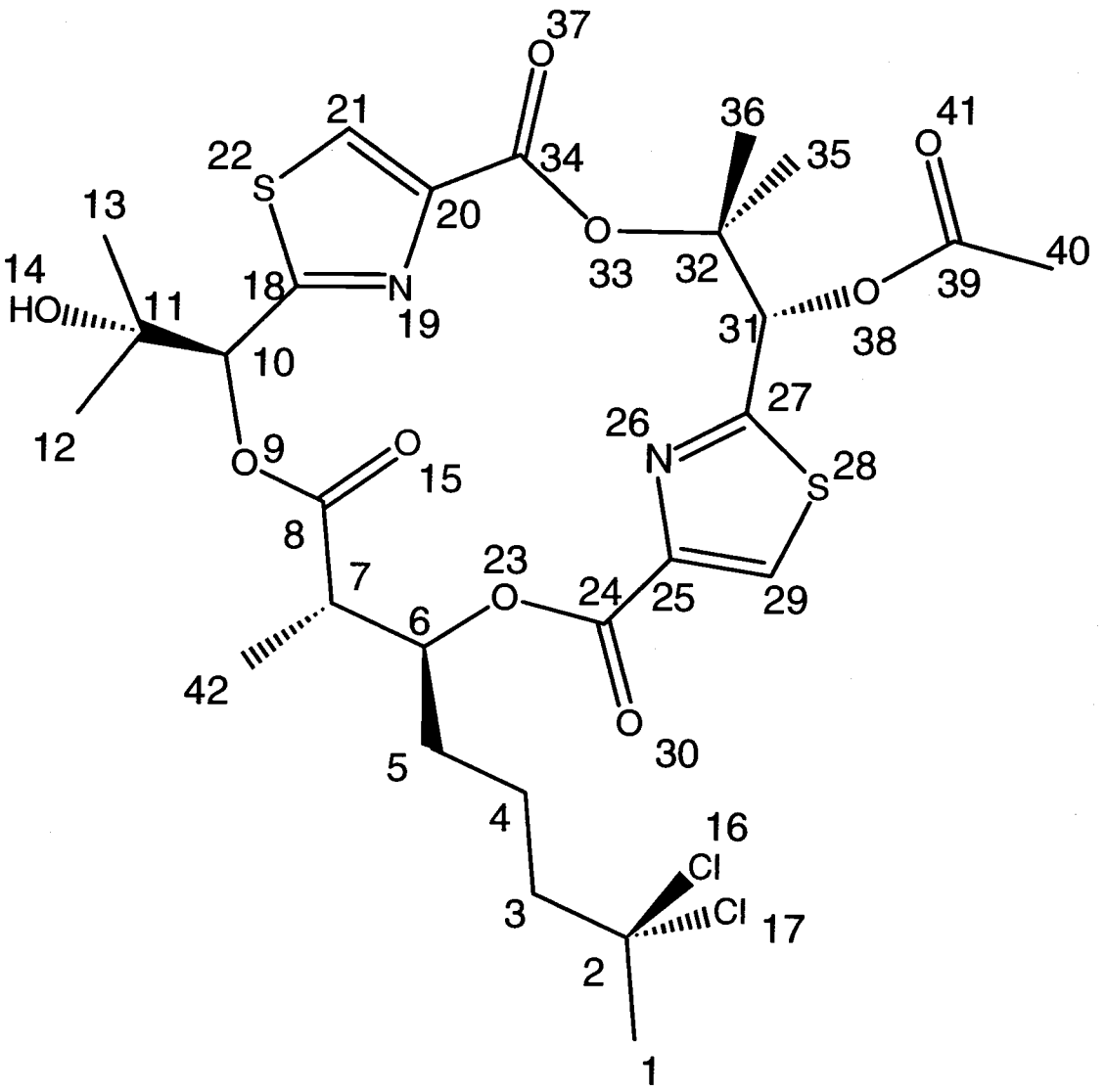


Figure 2.3 The absolute structure of hectochlorin. The single crystal structure of hectochlorin agrees with the structure determined from NMR and MS data. Structural features of note include four ester linkages, a saturated chlorinated carbon chain, and two thiazole rings. The numbering scheme is arbitrary.

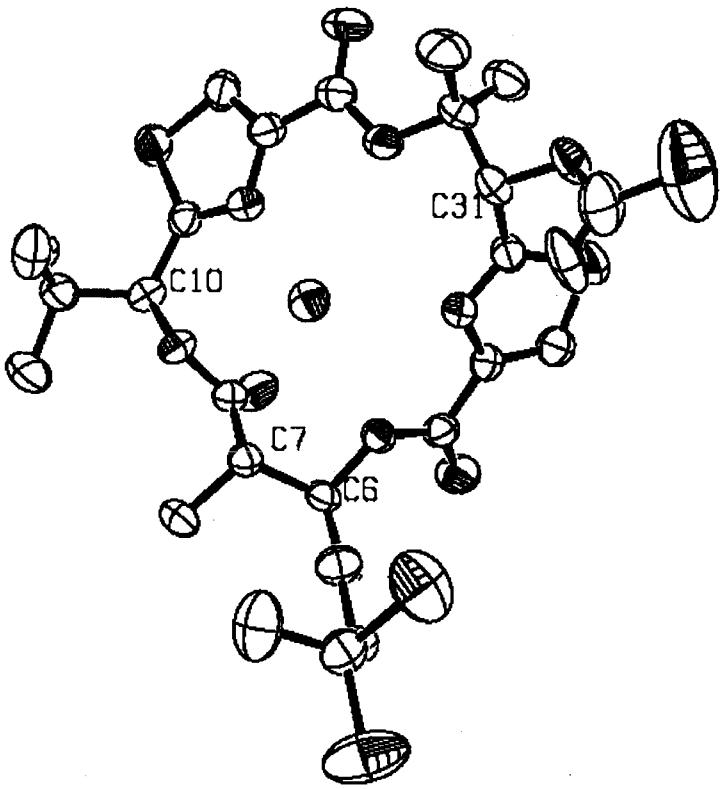


Figure 2.4 ORTEP representation of the non-hydrogen atoms of hectochlorin. The absolute structure is defined by x-ray crystallography and anomalous scattering data. Ellipses are drawn at 50 % probability. The detailed structure shows the stretching, rocking, and scissoring of bonds in the molecule.

the stereochemistry. Furthermore, to further study the value of **1** as a therapeutic agent or its industrial applications, large amounts of material are required. With the absolute configuration now defined, development of large scale synthesis and derivatization of the correct stereoisomer can now proceed.

The refined single crystal structure confirms the structure obtained earlier from NMR and MS data. The atoms in the hectochlorin crystal structure agree with the molecular formula, $C_{27}H_{34}N_2O_9S_2Cl_2$. The molecule contains two thiazole rings, four carbonyls, two carbons with geminal methyls, and a saturated carbon chain with geminal chlorine substituents. Additionally, a hydrogen bound solvent molecule is clearly evident in the crystal structure.

The crystal structure shows three ester linkages and two thiazole rings form a large ring of hydrogen bonding contacts to the solvent molecule. A single solvent molecule is encompassed by both imino nitrogens in the thiazole rings, and the hydroxyl group from a symmetry related molecule, which indicates the solvent molecule is constrained in the crystal by at least three hydrogen bonds. The N19 to solvent oxygen distance is 2.97 Å, the N26 to solvent oxygen distance is 2.80 Å, and the distance between the solvent oxygen and symmetry related O14 is 2.65 Å. The angle formed by N19, solvent oxygen, and N26 is 95.96°. The angle between N19, solvent oxygen, and symmetry related O14, is 110.78°. The N26, solvent oxygen, symmetry related O14 angle is 116.48°. Additionally, three ester oxygens are 3.08 Å, 3.19

Å, and 3.23 Å away from the solvent molecule. Small deformation of the structure may reduce these distances and present a second perimeter of hydrogen bond contacts. Hydrogen bonds effectively ensnare the solvent molecule within the macrocyclic ring.

The gross structure of **1** is similar to dolabellin **2**, shown in Figure 2.5. **2** is isolated from the sea hare*Dolabella auricularia*, and exhibits cytotoxic activity against HeLa-S_3 cells (16). **2** also has four stereocenters, with two centers differing in absolute stereochemistry from **1**. Differing biogenesis pathways may explain why the stereochemistry at C31 of **1** differs from the stereochemistry at the analogous stereocenter in **2**. That is, a different stereogenic precursor is selected in the biosynthesis of each molecule; glyceric acid is probably condensed into a thiazole ring in **2**, where α,β–dihydroxyisovalerate or dimethylallyl diphosphate is likely the precursor condensed in **1** to form a similar ring. The other stereocenter in **2** that differs from the analogous stereocenter in **1** is at C7. In **2**, the stereocenter in question was defined by synthesis of two possible fatty acid fragment isomers under stereoselective conditions. The isomers were characterized by optical rotation and NMR. The assignment was based on comparison of synthetic R, S and synthetic R, R to the relevant degradation product of **2**. Presuming no error in assignment from the optical data, the biosynthesis pathways must be different.

To further discuss the structural details and possible biogenesis pathways, consider the theoretical hydrolysis fragments of hydrolysis of the

1 **2**

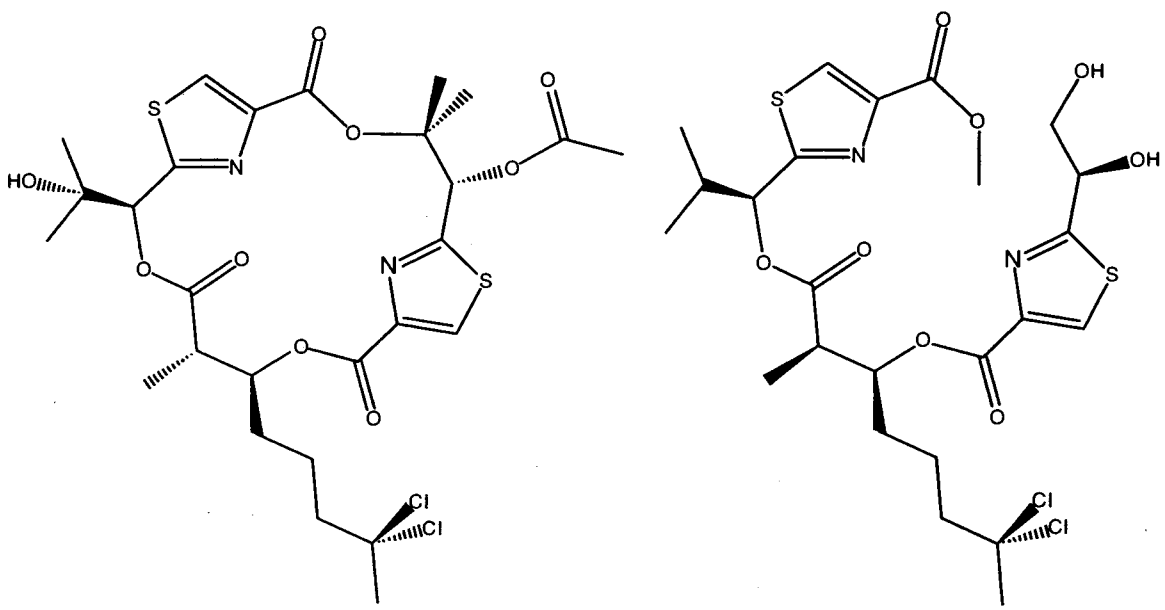


Figure 2.5 Structure comparison between hectochlorin, **1**, and dolabellin, **2**. Hectochlorin **1** is isolated from *Lyngbya majuscula*. **2** is isolated from the sea hare *Dolabella auricularia*. Inspection reveals a similarity between the gross structure of the two compounds. Furthermore, each compound has four stereogenic carbons.

esters would result in the fragments represented in Figure 2.6. The imaginary hydrolysis would yield two equivalents of fragment A, which contains a thiazole ring, the predominant chromophore in **1**. The biogenesis of this fragment can be hypothesized to proceed by at least two routes. One possibility is the condensation a cysteine amino acid to a dimethylallyl diphosphate. The dissociation of the PP_i creates a stable allylic carbocation that could be attacked by the nucleophilic amino group. Hydration of the double bond to form a secondary alcohol, oxidation to form the tertiary alcohol, and oxidation to form the thiazole ring are required steps to ultimately form fragment A. Another, more likely possibility is the condensation of cysteine with α,β–dihydroxyisovalerate, an intermediate in valine biosynthesis. Provided that cultured *Lyngbya majuscula* produces **1**, feeding studies can be proposed to distinguish between the various biosynthetic pathways. Enrichment of carbon 13 subsequent to feeding cultured *Lyngbya majuscula* labeled dimethylallyl diphosphate would support the hypothesis that dimethylallyl diphosphate is included in the biosynthesis. The second hypothesis would be supported by enrichment, subsequent to feeding cultured *Lyngbya majuscula* labeled α,β–dihydroxyisovalerate. The role of cysteine as a precursor could be probed with labeled cysteine feeding studies. Cellular extracts may be more appropriate to perform biosynthesis studies on, presuming the cellular uptake

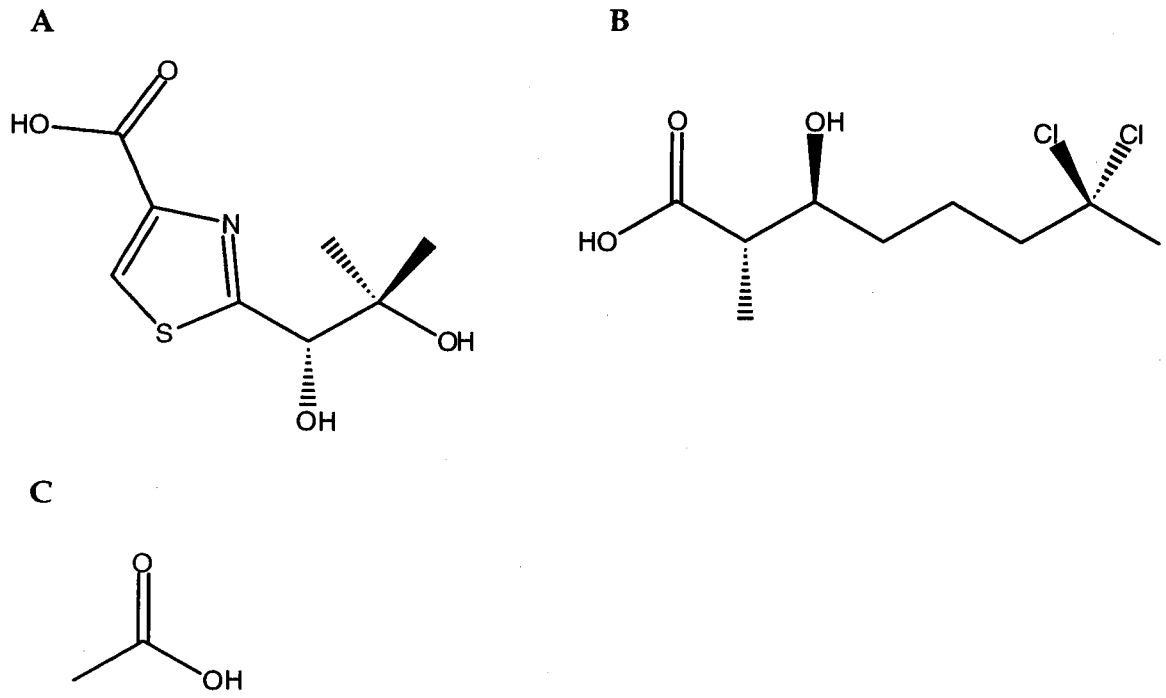


Figure 2.6 The fragments generated from a theoretical hydrolysis of hectochlorin. Two equivalents of fragment A would be generated. Fragment A contains a thiazole ring and a vicinal diol. Fragment B is an α -methyl-β-hydroxy-η-dichloro fatty acid. Fragment C is acetic acid.

of charged species, such as dimethylallyl diphosphate, and α,β–dihydroxyisovalerate is poor.

Fragment B is likely the product of polyketide anabolism. A likely path of biosynthesis is an acetyl-CoA starter unit condensed with three units of malonyl-CoA. The α-methyl moiety could result if the penultimate module lacks dehydration function, and instead oxidizes the polyketide to an enol, followed by stereospecific methylation, which results in an α-methyl-β-hydroxyl as found in **1**. An alternate possibility of an acetyl-CoA starter unit, condensed with two malonyl-CoA units, then finally a methylmalonyl-CoA unit, also explains the presence of the α-methyl moiety. This hypothesis requires the stereospecific loading and incorporation of methylmalonyl. Fragment C is very likely to be the result of a simple acetylation. Labeled acetate would probe this possibility.

SUMMARY

In this chapter, the absolute configuration of hectochlorin, as determined by x-ray crystallography, has been reported as 6S, 7S, 10S, 31S, and its possible biogenesis has been discussed. Marine natural products are a source of novel chemical backbones and provide skeletons for combinatorial synthesis, therapeutic agents, and insight into new biosynthesis possibilities. *Lyngbya majuscula* continues to yield secondary metabolites that are potentially valuable as therapeutic agents or have industrial applications.

Chapter 3 Crystallization of TAR

INTRODUCTION

A crystal structure can give a wealth of information for structure-assisted drug design. Obtaining diffraction quality crystals is the first step in the process. Reported in this chapter are the results of attempts to crystallize an RNA molecule of significant biological importance, the *trans*-acting response element of HIV, to advance structure-assisted drug design efforts.

Efficient transcription of the HIV genome requires interaction between the viral *trans*-activating transcription protein (Tat) and the *trans*-activating response element (TAR). TAR is the 59 nt leader sequence found at the 5'-end of every viral mRNA transcript. The sequence folds into a hairpin conformation, with a two to three base bulge, that is specifically recognized by Tat. The bulged hairpin structure is critical for the Tat-TAR interaction. This protein-RNA interaction results in increased RNA polymerase processivity. This interplay does not have an analogous host interaction, which makes it an attractive target for the development of anti-viral agents (17).

After incorporation into the host cell, the viral RNA genome is transcribed into DNA by viral reverse transcriptase (RT). The viral genome is then inserted into the host's DNA by viral integrase (IN). After a latency period the viral genome is replicated by the host transcription machinery, and as the viral messages are transcribed, the Tat-TAR interaction assures processivity and complete transcript production. HIV protease (PR) then

enables packaging of the viral particle by modification of capsid proteins (17). These four viral activities, RT, IN, Tat, and PR, are potential targets for drug design. Currently, the arsenal against HIV includes drugs that inhibit RT and PR. The sloppy copying of RT introduces mutations that result in resistant variants. Hence, the rapid mutation rate limits the long term potency of any one agent. Therefore, novel agents and strategies are in constant demand as resistant viruses are selected.

Tat is an 86 amino acid viral protein with five domains. The 48 residues at the N-terminus form a transcriptional cofactor binding domain. An arginine rich domain motif (ARM), from residue 49 to 58, is required for recognition of TAR. The other domains include a core domain and a cysteine rich region. The cysteine rich domain can complicate biochemical studies with Tat by forming non-native disulfide bridged complexes (18). Polyarginine polypeptides bind specifically and efficiently to TAR. However, polylysine polypeptides do not bind specifically and the binding constant is reduced by a factor of 10 relative to polyarginine. Arginine 52 and 53 of the ARM are the principal residues in the Tat-TAR interaction (19). A single arginine amino acid binds specifically, and will effectively compete with the entire Tat protein (20, 21). Arginine and argininamide are common substitutes for Tat in structural studies. The guanidinium group of the arginine residue forms an arginine fork, a network of hydrogen bonds, with the G26·C39 base pair in TAR (19).

The viral genome is laid out in tandem repeats. Without Tat present, the transcription complex of host machinery is formed, but has low processivity. The low processivity creates abortive complexes that can not lead to a new generation of infectious viral particles. In contrast, the Tat-TAR complex is processive and generates complete viral messages. The Tat-TAR complex alone can maintain the phosphorylation of the C-terminal domain of RNA polymerase II (22). Accessory proteins to Tat also increase processivity of the transcription machinery. After Tat binds to TAR, cyclin T associates with Tat and the TAR loop region. The associated kinase, CDK 9, increases processivity by hyper-phosphorylating the RNA polymerase machinery (23, 24, 25).

The native TAR sequence is 59 nt long, however, the consensus sequence required for activity is a truncated component of this. A uracil containing bulge, flanked by a stem and a stem loop, is conserved in the consensus sequence (26). Figure 3.1 depicts the secondary structure of the wild type and consensus sequences. The conserved uracil in the bulge is protected in biochemical assays, which suggests that it may be tucked into the major groove, perhaps forming a base triplet between U23 and the A27·U38 base pair (20, 27, 28).

The formation of the U23·(A27·U38) base triplet is an important feature to the binding of Tat. Mutation studies show that the loss of the base pair portion (A or U) of the trio deletes activity. Additionally, a mutation of the

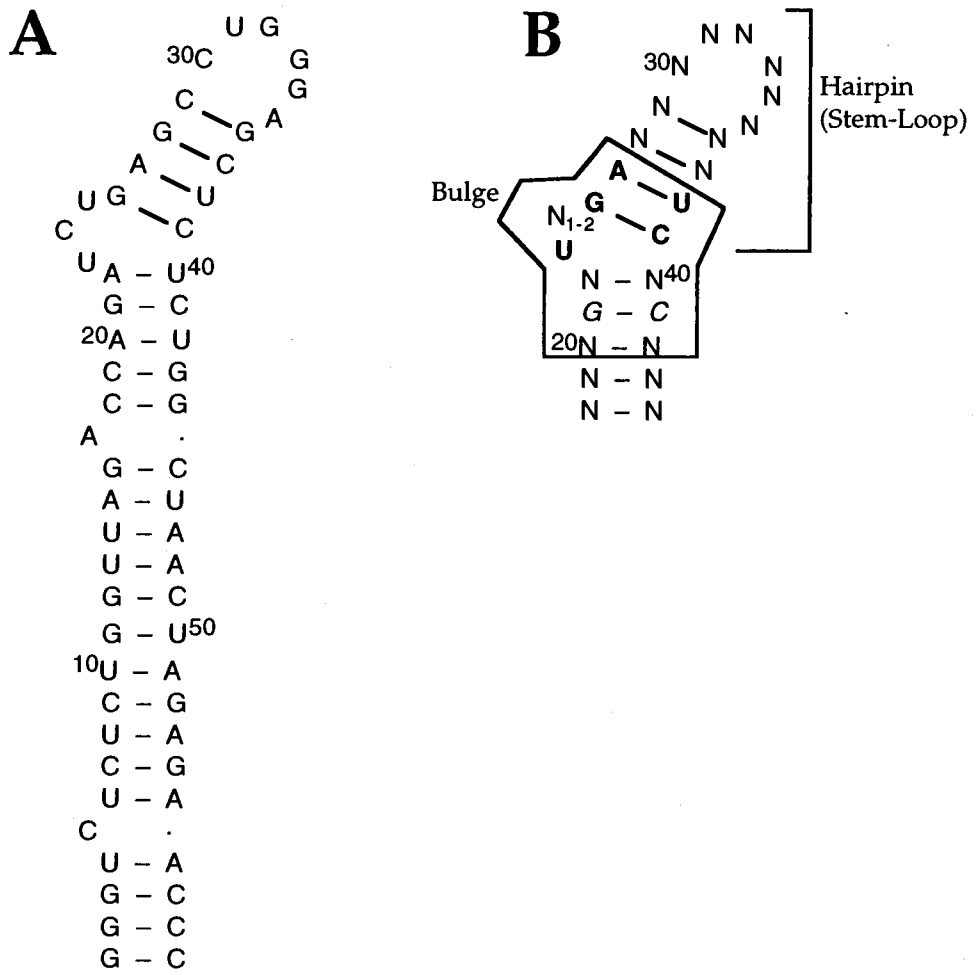


Figure 3.1 The wild-type sequence of HIV TAR and the consensus sequence of TAR. Panel A is the secondary structure of the 59 nucleotide long TAR sequence. The major features of the planar structure are a stem, bulge, and stem loop. Panel B shows the conserved structure required for Tat recognition. The required residues are in bold, and elements with a moderate effect on binding are italicized. The G26·C39 base pair may be required to form the arginine fork interaction with Tat. U23 in the bulge may interact with base pair A27·U38 to form a triplet (20, 27). supporting the requirement of the base triplet (20). The base triplet may stabilize a widening of the major groove of the helix that facilitates binding of Tat (29, 30).

bulge residue destroys activity. However, a triple mutation that restores a base pair (G·C) and places a C in the bulge is recognized and bound by Tat,

The structures of TAR unbound and bound to Tat are critical to understanding this important step in viral replication. Previously reported structures of TAR present conflicting results and leave unanswered questions as to what are the detailed three dimensional structures of the bound and unbound forms of TAR.

Puglisi *et al.* (1992) used NMR to investigate the structure of TAR. Their construct was a bulged, stem loop. The model of TAR bound to arginine resulting from their data suggested that arginine forms an arginine fork with the G26·C39 base pair. Their model also shows that U23 is positioned in the major groove in such a way as to form a base triplet with the A27·U38 base pair (31). Aboul-ela *et al.* (1995) also used NMR to investigate the TAR hairpin structure. With a larger number of data to restrain their model building, which should lead to a more accurate model, they proposed a different structure. Similar to Puglisi *et al.*, their model contains an arginine fork with G26·C39 and U23 is tucked into the major groove of the helix. However, the distance between U23 and its potential partners is too great to form a base triplet (32). This leaves us with two conflicting perspectives of the TAR structure, both models are from NMR, a technique that relies on close contacts for developing distance constraints.

A third model has been presented by Ippolito and Steitz (1998) who solved the crystal structure of a bulged duplex TAR construct in the presence

of Ca^{+2}. Their data gives yet another model for TAR. This construct lacked the loop portion of the consensus sequence. One advantage of a crystal structure is that the solvent environment can be defined. In this model we see that U23 is displaced out of the major groove, nowhere near the A27·U38 base pair. The bulge conformation was stabilized by calcium ions (33). The position of U23 in this model is inconsistent with biochemical evidence and both NMR models, which suggest U23 is protected in the major groove. Low concentrations of magnesium have been shown to change the conformation of TAR, thus more information is required to determine if this crystal structure is a relevant unbound conformation (34).

There are three structures of TAR, and each tells a different tale. There is evidence that a conformational change occurs upon binding arginine, so each may be a relevant piece of a dynamic process. Long and Crothers suggest the unbound conformation of TAR is bent at the bulge, and upon binding, the stems stack co-axially (27, 28). Clearly, there is a need for detailed crystal structures of arginine or argininamide bound TAR, and the unbound form of TAR. The structures will support or challenge the presence of a base triplet and the conformation change that may occur upon ligand binding.

Detailed models are key to structure assisted drug design. A crystal structure can show details of the surrounding solvent, which can give insights into drug design and improvement. All available structures should be considered in designing cognate contacts for the unbound conformation and contacts that will increase stabilization of the bound form.

The first step in solving a crystal structure involves determining the conditions that generate diffraction quality crystals of a biologically relevant construct. Various salts, buffers, and precipitants can be screened to find the optimal conditions that foster crystal growth. Large polyelectrolytes can be problematic; nucleic acids are, more or less, a uniform landscape of repelling negative charges that can shift slightly and distort a crystal lattice. This distortion reduces the quality of the diffraction data and hence, the quality of the model fit to the data.

To improve lattice contacts in nucleic acid crystals, Ferre'-D'Amare' *et al.*, incorporated non-native structure motifs that lead to the crystallization of several large RNA molecules. Originally observed as a lattice contact, the GNRA tetraloop (TL), and GNRA (where G is guanine, N is any nucleotide, R is a purine, and A is Adenosine) tetraloop receptor (TLR), are observed in the P4-P6 domain of the group I intron ribozyme and the hammerhead ribozyme. Unbound, the adenosine bases in the TLR are stable in an adenosine zipper motif. The bound conformation of the TL is stabilized by a base quartet, base pairing, and an adenosine platform (35). A TL and TLR were successfully employed as crystallization modules, an inter-domain stabilizing interaction, for the crystallization of the group II intron 5-6 domains, and the hepatitis delta virus ribozyme (36, 37).

With this strategy in mind, the sequences in Figure 3.2 were developed to utilize this interaction. The binding of one TL into a TLR on another molecule should facilitate crystallization by providing a defined interaction.

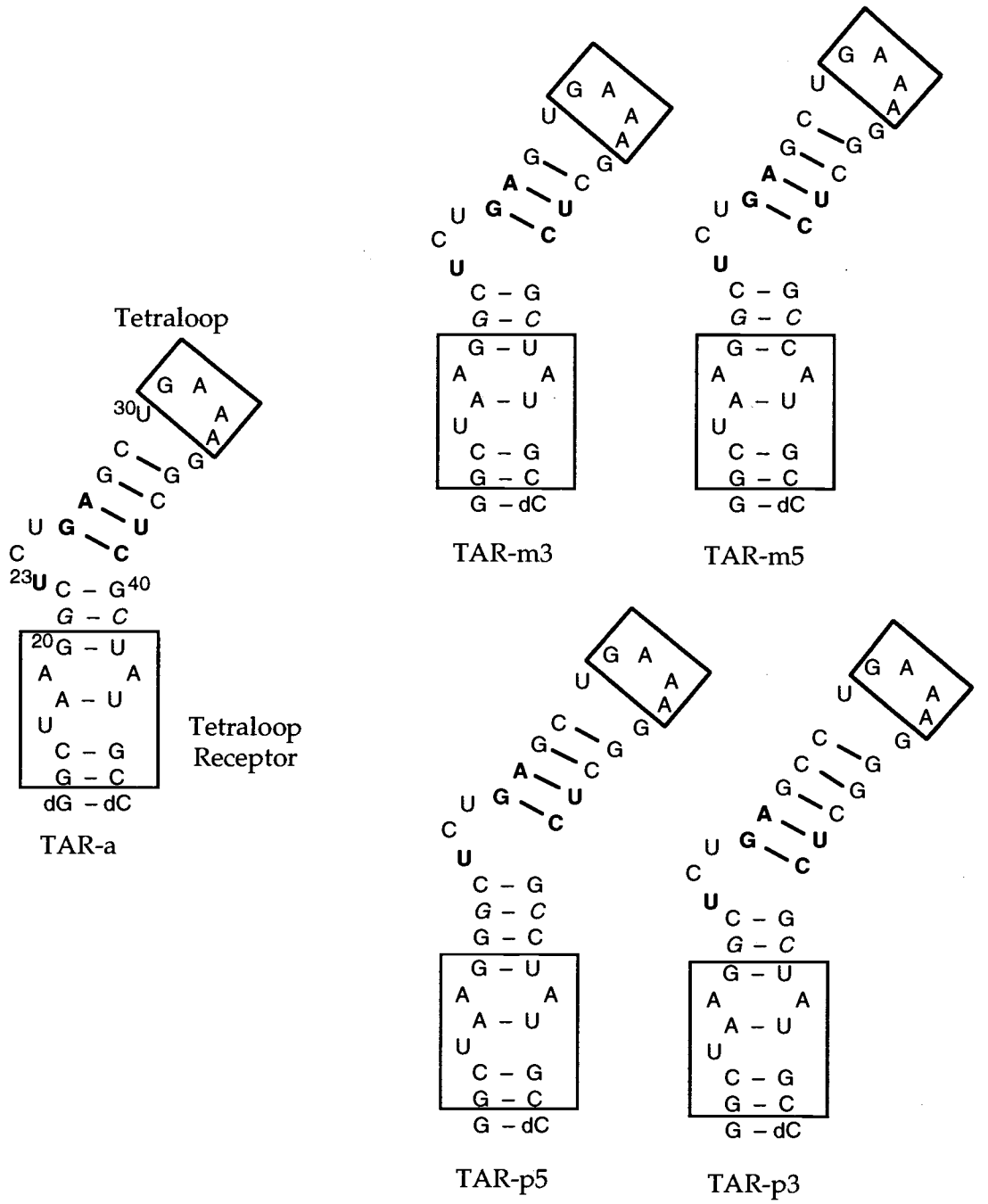


Figure 3.2 TAR sequences used in this study. The sequence and proposed secondary structure of TAR-a and derivatives. These variants of TAR-a differ in phase and spacing. The conserved bases are in bold, and bases with a moderate effect on Tat binding are italicized. The tetraloop and tetraloop receptor regions are marked with boxes.

As shown by the wild type and consensus sequences in Figure 3.1 the binding of Tat to TAR is unaffected by residue substitution in the loop region of TAR. Also, the stem region distal to the loop allows broad substitution. The sequence and positioning of the TL and TLR in TAR constructs at the loop and on the stem leaves the consensus sequence pristine. The polarity and spacing of the TLRs differs between the designed sequences to provide various contact orientations between the modules. The TL - TLR constructs should facilitate crystallization and ultimately provide biologically relevant structures. Indeed, the constructs did prove to be crystallogenic in a variety of conditions.

MATERIALS AND METHODS

TAR sequences were synthesized on an Applied Biosystems Inc. DNA synthesizer. NAP-10 columns of G-25 Sephadex, DNA grade, were obtained from Amersham Pharmacia. Reagent grade concentrated ammonium hydroxide was obtained from J.T. Baker. Absolute ethanol, 1 M tetrabutylammonium fluoride (TBAF) in tetrahydrofuran (THF), 99.8% Tris-HCl and 99.8%+ grade ammonium acetate were obtained from Aldrich. 3 cc syringes and multiwell tissue culture plates were obtained from Becton Dickinson & Co. HPLC grade acetonitrile, Eppendorf tubes, sterile 50 mL and 15 mL conical tubes were obtained from Fisher Scientific. Semi-preparative (10 mm x 250 mm) C-18 Absorbosphere HPLC column was obtained from Alltech. 0.45 μm Nylaflow 47 mm nylon membrane filters obtained from

Gelman, and 0.45 μm Durapore membrane filters 47 mm were obtained from Millipore. pH determinations were performed on a Corning 120 pH meter, that was calibrated prior to measurement. Natrix screen solutions, and silanized glass cover slips were purchased from Hampton Research. The HPLC system was composed of a Hewlett-Packard Ti 1050 pump and Hewlett-Packard 1040A UV detector. 10X triethylammonium acetate (TEAA) stock aqueous solvent was made up using 700 mL autoclaved distilled deionized (dd) water, 75.25 mL concentrated acetic acid (Sigma), and 139.4 mL ACS grade triethylamine (Sigma) titrated with acetic acid to pH 6.5. RNA concentrations were determined using a Hewlett-Packard 8452A spectrophotometer and extinction coefficients in Table 3.1.

Removal of sequence from synthesis cartridge and deprotection of amines

2 mL of 3 : 1 ammonium hydroxide : ethanol solution is loaded into a 3 cc syringe, purged to remove trapped air and then attached to one end of the synthesis cartridge. A second 3 cc syringe is placed on the other end with 0.5 mL of head space. The syringes are alternately pushed/pulled 5 times slowly to drive the solution through the cartridge resin. The agitation is repeated after 15 min. then every 30 min. for 3 hours. Between agitation steps the solution is positioned to bathe the cartridge resin. Finally, the solution is expelled into glass vials, sealed and heated 12 hours at 55° C.

Table 3.1 Molecular weights and extinction coefficients at 260 nm used for TAR constructs. This table lists the extinction coefficients of the TAR constructs used in this study. The extinction coefficients, molar^{-1} cm^{-1}, were calculated at 260 nm using neighboring interactions. Masses were calculated with ammonium counter ion.

TAR construct	ε260 nm (M^{-1} cm^{-1})	Molecular Weight (g/mol)
TAR-a	323,900	10,994.8
TAR-p5	341,000	12,273.2
TAR-m3	308,600	10,904.4

Deprotection of 2' hydroxyls

The material is transferred to Eppendorf tubes using two 100 µL washes to rinse vials and assist transfer. The sample is then lyophilized. The dry powder is dissolved in 300 µL of 1 M TBAF in THF and gently agitated 16 hours.

NAP-10 Column Purification

NAP-10 columns are prepared by draining until mobile phase reached the G-25 resin bed. TBAF reaction products of like sequences are combined and transferred to the column using 200 µL dd water rinses (total volume 1 mL). Material is eluted by loading 800 µL dd water, followed by up to six 500 µL aliquots of dd water. The fractions are then ethanol precipitated by the addition of 50 µL 8 M ammonium acetate, 1100 µL absolute ethanol, chilling at -80°C for 25 min., then spinning 14,000 rpm, 25 min. at 4°C. Fractions that contained a pellet are combined and ethanol precipitated again.

HPLC purification

Table 3.2 illustrates the gradient used in the separation. The (A) solvent is 0.1 M TEAA pH 6.5 and the (B) solvent is acetonitrile. The solvents are filtered with 0.45µm and degassed with helium approximately 15 min. prior to use. Samples are centrifuged prior to injection onto the HPLC

Table 3.2 HPLC gradient for RNA purification. RNA sequences are purified by C-18 high performance liquid chromatography using 0.1 M Triethylammonium acetate (pH 6.5), (A), and HPLC grade acetonitrile, (B), in the following gradient. Solutions were filtered with 0.45 μm membrane filters and degassed with helium prior to use. The flow rate was 2 mL/ min. for the entire gradient.

Time (min.)	Mobile Phase (%B)
0	0
5	0
10	20
20	30
25	30
27	50
30	50
35	0

column. The flow rate is 2 mL/min. Elution is monitored by UV. Fractions are dried down, combined by sequence using 100 mM Tris-HCl pH 7.5, then concentrated by ethanol precipitation.

Detritylation

Sequences are detritylated in the following manner: to 50 µL of sample in 100 mM Tris-HCl pH 7.5 an equal volume of glacial acetic acid is added. The reaction is allowed to run for 30 min. at room temperature. The reaction is quenched by the addition of 10 µL 1 M Tris-HCl pH 7.5. 10 µL 8 M ammonium acetate, 80 µL dd water and 600 µL absolute ethanol are then added for precipitation.

Hanging drop crystallization setups

The concentration of each purified detritylated sequence is determined spectrophotometrically using the extinction coefficients in Table 3.1. The sample is then concentrated with an ethanol precipitation and brought up to a concentration of 4 mg/ mL. Samples are centrifuged prior to setup. Equal volumes of each sequence and Natrix screen solution, plus or minus argininamide, are dispensed onto a silanized glass cover slip as per Table 3.3. The cover slip is then inverted and placed over a 750 µL reservoir with a greased lip.

Table 3.3. Crystallization conditions for hanging drop setups of TAR constructs. Crystals formed in 1 to 4 weeks. Precipitant concentrations are reported in percent volume/volume. Final concentrations are reported and the total drop volume was approximately 20 μL. Both 2-methyl-2,4-pentanediol (MPD), and polyethylene glycol (PEG) were used as precipitant this study. Argininamide ($ArgNH_2$) was used as a Tat analogue.

Sample, Concentration	Buffer	Salts, Other	Precipitant (% (v/v))
TAR-m3, 1.78 mg/mL	22 mM Tris-HCl (pH 7.5)	1 mM $ArgNH_2$, 22 mM NH_4OAc, 4.4 mM $MgCl_2$, 4.4 % MPD	10 % MPD
TAR-p5, 2 mg/mL	25 mM NaCacadylate (pH 6.5)	40 mM $MgOAc_2$, 7.5 % PEG 400	15% PEG 400
TAR-a, 1.78 mg/mL	22 mM Tris-HCl (pH 7.5)	1 mM $ArgNH_2$, 22 mM NH_4OAc, 4.4 mM $MgCl_2$, 4.4 % MPD	10 % MPD
TAR-a, 2 mg/mL	25 mM Tris-HCl (pH 7.5)	1 mM $ArgNH_2$, 22 mM NH_4OAc, 4.4 mM $MgCl_2$, 4.4 % MPD	10 % MPD

Freezing Conditions

Crystals are mounted in a thin nylon loop, and passed through cryoprotectant quickly to exchange the mother liquor. The cryoprotectant is composed of a solution of salts and buffer at a concentration equal to initial concentration crystallization conditions, plus additional 2-methyl-2,4-pentanediol to a final of 20% (v/v). The crystal is submersed in liquid nitrogen and transferred to the goniometer with cryotongs. Data are collected at -180°C using a Rigaku RU-300H rotating anode x-ray generator and an RAXIS IV image plate detector.

RESULTS AND DISCUSSION

The modular TL-TLR approach proved to be successful, providing a handful of conditions that give rise to crystal growth (Table 3.3). Crystallizing conditions have proven reproducible for TAR-a sequences (Figure 3.3). Panel A of Figure 3.3 shows a colorless crystal, approximately 0.1 x 0.1 x 0.1 mm^3, from a well containing TAR-a and argininamide. Panels B and C of Figure 3.3 present crystals from wells containing the TAR-a sequence, both crystals are colorless, and approximately 0.2 x 0.2 x 0.2 mm^3. Second generation sequences, TAR-p5 and TAR-m3, proved crystallogenic as well. Panels D and E of Figure 3.3, show crystals grown from wells containing TAR-m3 with argininamide, and TAR-p5, respectively.

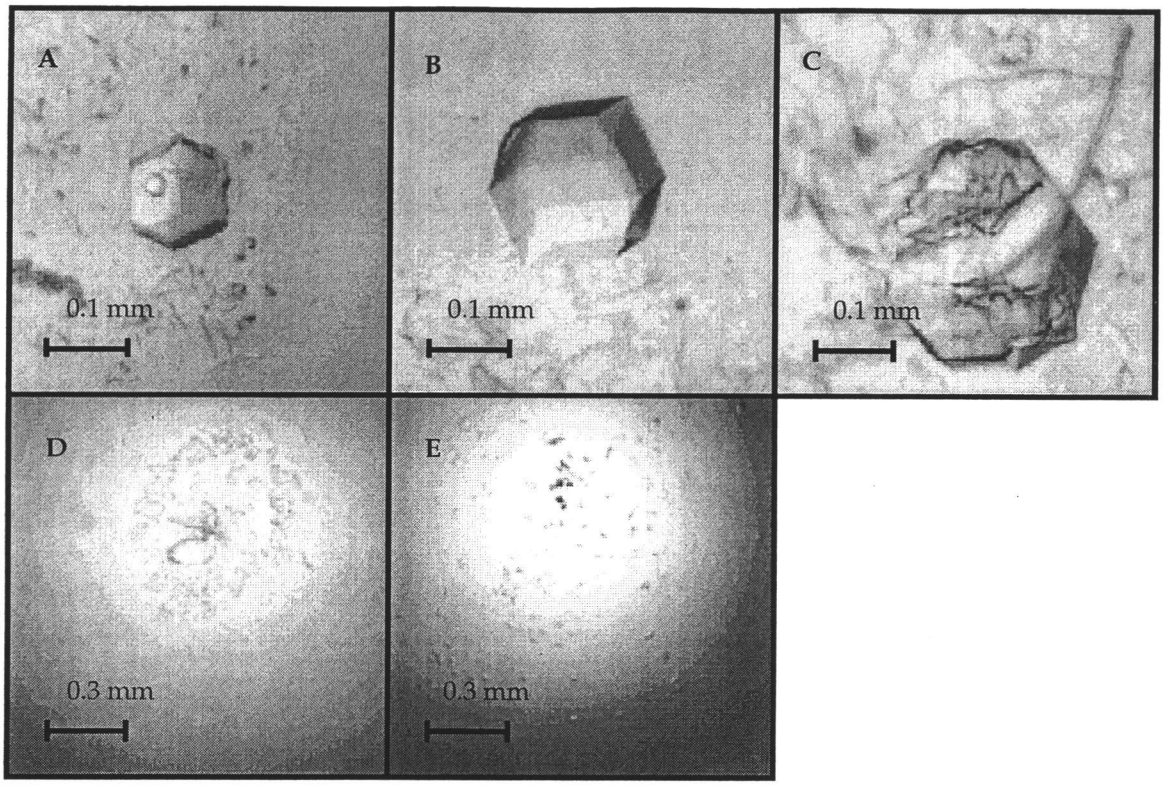


Figure 3.3 Crystals grown from wells containing TAR sequences. Panel A shows a colorless, 0.1 x 0.1 x 0.1 mm^3, crystal grown from a well containing TAR-a and argininamide. Panels B and C are pictures of colorless 0.2 x 0.2 x 0.2 mm^3 crystals grown from wells contain TAR-a. Panel D shows two football shaped, 0.2 x 0.1 x 0.15 mm^3, crystals grown from wells containing TAR-m3 and argininamide. Panel E shows a shower of cubic crystals grown from a well containing TAR-p5.

The crystals were checked for their diffraction quality, see Figure 3.4. The TAR-a crystal in Figure 3.3, panel A, was screened, and diffracted out to about 12 Å (Figure 3.4, Panel A). The TAR-m3 crystal shown in Panel D of Figure 3.3, diffracted out to approximately 15 Å (Figure 3.4, Panel B). The resolution of this data is poor, and cannot be used to solve the structure. Unfortunately, reliable unit cell dimensions and space group could not determined from the data collected.

One common feature of all the reported crystallization conditions is magnesium, a divalent cation. This is not unexpected, as cations are often a requirement for the folding of native RNA structures, such as ribozymes. The freezing conditions for storage and protection from degradation in the x-ray beam have been determined. This will facilitate transportation and data collection at synchrotron sources. A synchrotron beam source may be a requirement for collection of higher quality data from these crystals.

SUMMARY

Conditions have been determined that produce crystals of a biologically relevant sequence of TAR. Crystals of TAR and TAR bound to argininamide have been generated. Furthermore, crystallization conditions may be refined to produce larger and higher quality crystals. Simplification of the phase problem utilizing MAD may be achieved with brominated uridine derivative sequences of TAR. A quality crystal that diffracts to high resolution can be used to refine an accurate models of TAR for drug design purposes.

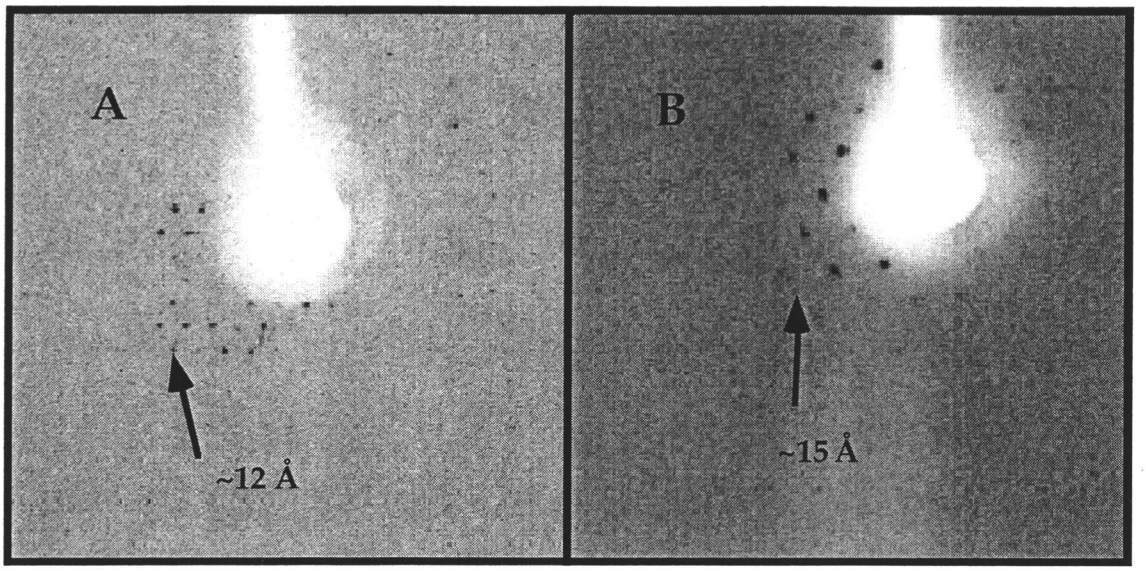


Figure 3.4 Diffraction patterns from TAR crystals. Panel A is a diffraction pattern collected on a crystal grown from TAR-a and argininamide. Reflections were observed as far as 12 Å. Panel B is a diffraction pattern collected on a crystal grown from TAR-m3 and argininamide. Reflections only out to ~ 15 Å were observed in this data.

Chapter 4 Conclusion

The field of structure-assisted drug design continues to mature. As the number of receptor and ligand structures increases, the ability to tailor molecules for a specific purpose from structural information alone becomes more tangible. The further investigation of hectochlorin, and possible development into a role in industry, or as a therapeutic agent can begin. Determination of the absolute configuration is an early step in its development. More biochemical information, such as the specific receptors it antagonizes or agonizes, is required to complete further steps towards maturation. The molecule can be made available for bench-top and computational screens. As novel assays evolve and new applications are discovered, the richness of structure libraries will be fully appreciated. Aside from biological activity, it is worthwhile to survey the scope of chemical compounds produced by secondary metabolite rich organisms such as *Lyngbya majuscula*. Understanding the complete range of chemistry available to the organism may lead to the production of novel products through genetic manipulation.

We have made steps towards determining the three-dimensional structure of TAR. At least two different rational drug design approaches have begun development and could be assisted by an accurate model of TAR. A split and mix technique has identified a peptoid analogue of Tat as a potent inhibitor that blocks TAR·Tat formation at nanomolar concentrations (38). Furthermore, aminoglycosides with arginine analogue substituents have

been created that bind specifically to TAR (39). Computer modeling of these Tat analogue compounds with TAR may lead to improvements in both of classes of drugs.

Both the determination of the absolute structure of hectochlorin, and results from TAR crystallization attempts, demonstrate structure-assisted drug design is taking strides toward maturation.

BIBLIOGRAPHY

1. D. E. Hammerschmidt, *The Journal of Laboratory and Clinical Medicine* **132**, 556 (1998).

2. P. Y. S. Lam, et al., *Science* **263**, 380-384 (1994).

3. A. Wlodawer, J. Vondrasek, *Annual Review Biophysics and Biomolecular Structure* **27**, 249-284 (1998).

4. S. Wang, et al., *Journal of Medicinal Chemistry* **39**, 2047-2054 (1996).

5. R. L. DesJarlais, et al., *Proceedings of the National Academy of Sciences* **87**, 6644-6648 (1990).

6. T. N. Helm, J. W. E. Dijkstra, R. F. Jr., S. Glanz, *American Family Physician* **43**, 908-912 (1991).

7. I. Amato, *Science* **262**, 32-33 (1993).

8. H. J. Smith, Ed., *Smith and Williams' Introduction to the Principles of Drug design and action* (Harwood academic publishers, 1998).

9. R. B. Herbert, *The Biosynthesis of Secondary Metabolites* (Chapman and Hall, ed. 2nd, 1989).

10. W. H. Gerwick, et al., *Journal of Organic Chemistry* **59**, 1243-1245 (1994).

11. L. T. Tan, et al., *The Journal of Organic Chemistry* **65**, 419-425 (2000).

12. K. E. vanHolde, W. C. Johnson, P. S. Ho, *Principles of Physical Biochemistry* (Prentice Hall, ed. 1st, 1998).

13. J. Drenth, *Principles of Protein X-Ray crystallography* (Springer, ed. 2nd, 1999).

14. G. M. Sheldrick (1997) SHELX-97-2, crystallographic program. University of Gottingen, Germany.

15. H. D. Flack, G. Bernardinelli, *Acta Crystallographica Section A* **A55**, 908-915 (1999).

16. H. Sone, et al., *Journal of Organic Chemistry* **60**, 4774-4781 (1995).

17. M. R. Adams, B. M. Peterlin, in *Molecular Biology and Biotechnology* R. A. Meyers, Ed. (VCH publishers, 1995) pp. 10-15.

18. K. M. Weeks, D. M. Crothers, *Biochemistry* **31**, 10281-10287 (1992).

19. B. J. Calnan, B. Tidor, S. Biancalana, D. Hudson, A. D. Frankel, *Science* **252**, 1167-1171 (1991).

20. J. D. Puglisi, L. Chen, A. D. Frankel, J. R. Williamson, *Proceedings of the National Academy of Sciences* **90**, 3680-3684 (1993).

21. J. Tao, A. D. Frankel, *Proceedings of the National Academy of Science* **89**, 2723-2726 (1992).

22. N. F. Marshall, G. K. Dahmaus, M. E. Dahmus, *The Journal of Biological Chemistry* **273**, 31726-31730 (1998).

23. B. R. Cullen, *Cell* **93**, 685-692 (1998).

24. C. Isel, J. Karn, *Journal of Molecular Biology* **290**, 929-941 (1999).

25. J. Karn, *Journal of Molecular Biology* **293**, 235-254 (1999).

26. S. Roy, U. Delling, C.-H. Chen, C. A. Rosen, N. Sonenberg, *Genes and Development* **4**, 1365-1373 (1990).

27. K. S. Long, D. M. Crothers, *Biochemistry* **38**, 10059-10069 (1999).

28. A. S. Brodsky, J. R. Williamson, *Journal of Molecular Biology* **267**, 624-639 (1997).

29. T. Hermann, E. Westhof, *Chemistry and Biology* **6**, R335-R343 (1999).

30. K. M. Weeks, D. M. Crothers, *Cell* **66**, 577-588 (1991).

31. J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, *Science* **257**, 76-80 (1992).

32. F. Aboul-ela, J. Karn, G. Varani, *Journal of Molecular Biology* **253**, 313-332 (1995).

33. J. A. Ippolito, T. A. Steitz, *Proceedings of the National Academy of Sciences* **95**, 9819-9824 (1998).

34. M. Zacharias, P. J. Hagerman, *Proceedings of the National Academy of Sciences* **92**, 6052-6056 (1995).

35. G. L. Conn, D. E. Draper, *Current Opinion in Structural Biology* **8**, 278-285 (1998).

36. A. R. Ferre-D'Amare, K. Zhou, J. A. Doudna, *Journal of Molecular Biology* **279**, 621-631 (1998).

37. A. R. Ferre-D'Amare, J. A. Doudna, *Journal of Molecular Biology* **295**, 541-556 (2000).

38. F. Hamy, et al., *Proceedings of the National Academy of Sciences* **94**, 3548-3553 (1997).

39. A. Litovchick, A. G. Evdokimov, A. Lapidot, *Biochemistry* **39**, 2838-2852 (2000).