

**Crystallographic and modeling studies
of intermolecular interactions of biological interest**

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

vorgelegt von
Eftichia Alexopoulos
aus Athen, Griechenland

Göttingen 2004

D7

Referent: Prof. Sheldrick Ph.D

Korreferent: Prof. Dr. Usón-Finkenzeller

Tag der mündlichen Prüfung: 29.04.2004

CONTENTS

1 INTRODUCTION	1
2 THEORETICAL BACKGROUND	3
2.1 Crystallographic methods	3
2.1.1 X-ray diffraction by crystals	3
2.1.2 Solution of the phase problem	5
2.1.3 Molecular replacement	14
2.1.4 Density modification	16
2.1.5 Structure refinement	17
2.2 Molecular modeling	19
2.2.1 Docking with FlexX	19
2.2.2 Energy minimization with Discover	23
3 THE 7-AMINOACTINOMYCIN D COMPLEX WITH NON-COMPLEMENTARY DNA	26
3.1 Introduction	26
3.2 Former work	29
3.3 Crystallization	30
3.3.1 The hexagonal crystals	30
3.3.2 The orthorhombic crystals	31
3.4 The orthorhombic form	32
3.4.1 Data collection	32
3.4.2 Structure solution	34
3.4.3 Structure refinement	37
3.5 The hexagonal form	39
3.5.1 Data collection	39
3.5.2 Structure solution	41
3.5.3 Structure refinement	46
3.6 Structure of the 7-AAMD complex	48
3.6.1 Overall structure	48
3.6.2 Conformation	52

3.6.3 Interaction with the 7-AAMD	54
3.6.4 Comparison with the hexagonal structure	55
3.6.5 Pseudosymmetry	56
4 THE D,L-ALTERNATING PEPTIDE H-(L-TYR-D-TYR)₄-LYS-OH	58
4.1 Introduction	58
4.2 Crystallization	59
4.3 Data collection	60
4.3.1 Native data	60
4.3.2 Iodide derivative data	61
4.4 Structure solution	61
4.4.1 Direct methods	62
4.4.2 SAD	62
4.4.3 SIRAS	63
4.5 Structure refinement	64
4.6 Structure of the H-(L-Tyr-D-Tyr)₄-Lys-OH	66
5 MOLECULAR MODELING STUDIES ON HERα LBD WITH ORGANIC LIGANDS	71
5.1 Introduction	71
5.1.1 Aim of the project	71
5.1.2 The human estrogen receptor alpha ligand binding domain	72
5.1.3 Working plan	78
5.2 Methods	78
5.2.1 Modeling	78
5.2.2 The ligands	79
5.3 Results and discussion	80
5.3.1 Modeling studies of the LBD with known agonists	80
5.3.2 Modeling studies of the ligands with the LBD	82
5.3.3 Modeling studies of mutated LBDs with the ligands	87

6 SUMMARY AND CONCLUSIONS	97
6.1 The 7-AAMD-ssDNA complex	97
6.2 The d,l-alternating peptide	100
6.3 Modeling studies on hERα LBD	100
7 REFERENCES	101
8 APPENDIX	106

LIST OF ABBREVIATIONS

A	adenine
7-AAMD	7-aminoactinomycin D
AMD	actinomycin D
BRU	bromouracil
C	cytosine
CC	correlation coefficient
DNA	deoxyribonucleic acid
<i>et al.</i>	<i>et alii</i>
F(cal)	calculated structure factor
F(obs)	observed structure factor
Fig.	figure
FOM	figure of merit
G	guanine
g^+	gauche +
g^-	gauche -
hER α	human estrogen receptor alpha
hER β	human estrogen receptor beta
L.S.	least squares
MAD	multiple anomalous dispersion
MIR	multiple isomorphous replacement
MR	molecular replacement
NCS	non-crystallographic symmetry
PDB	Protein Data Bank
r.m.s.	root mean square
SAD	single anomalous dispersion
SIR	single isomorphous replacement
T	thymine
Tab.	Table

FIGURES

Figures containing molecular graphics were created with the program Raster3D (Merrit & Bacon 1997) after preparation with the program DINO (A. Philippsen, <http://www.dino3d.org>).

Fig. 2.9 was taken from J. Drenth, *Principles of Protein X-ray Crystallography*, Sec. Edition, Springer Verlag, New York Berlin Heidelberg, p.225.

Figures 2.10, 2.11, 2.13 and 2.14 were taken from Rarey et al. 1996.

Fig. 5.3 was created with ViewerPro 4.2 (<http://www.accelrys.com>).

Fig. 5.4 was created with LigPlot (Wallace et al. 1995).

Auch für Harald

INTRODUCTION

The process of drug design is one of the most important tasks of pharmaceutical research. It involves both structural methods, like X-ray crystallography or NMR, as well as computer-based techniques, such as molecular modeling.

The importance of structural information can be explained by the key-lock principle. Most pharmaceuticals act as agonists or antagonists (key) for specific receptors (lock) in the human organism. Receptors are in general macromolecules, they can be proteins or enzymes or even DNA. Agonists are mostly small organic molecules that bind to these macromolecules and cause an intrinsic effect, while antagonists inhibit either reversibly (competitive inhibitors) or irreversibly the binding of a natural ligand to the receptor, causing a repression of the biological function. In order to develop a drug that will act either as an agonist or an antagonist for a specific target macromolecule it is extremely helpful to know where on the receptor it will act and how. The knowledge of the three-dimensional structure of the target molecule can supply important information about the binding site. The structure of the receptor complexed with its natural ligand is even more helpful as it provides the exact interaction points and types. Based on this information new compounds can be designed to specifically interact with the macromolecule.

The design of such lead compounds is more efficient if they are first tested *in silico* as regards their binding to the target receptor. This is the concept of molecular modeling studies, which have become quite popular in the last years, as a consequence of the development of ever faster computers. In this theoretical approach, the number of possible structures can be reduced by far, so the further organic synthesis is limited to only a few candidates. Nevertheless, the description of a biological system becomes more complex the bigger the degrees of freedom. While the energy calculation for small molecules is easy and reliable, it is almost impossible to predict the secondary structure of an unknown protein. However, the number of entries in the Protein Data Bank during the last decades has increased exponentially, so if the structure is not known yet, there is still the possibility of homology

modeling. But, even if the three-dimensional structure is given, it is still difficult and in many cases risky to predict its time-dependent conformational changes.

Structures obtained from X-ray crystallography on the other hand are very reliable, but the method is extremely time-consuming. The protein has to be expressed and purified before the actual crystallization can start. All these three processes present bottlenecks for protein crystallography and can each last from a few months to years.

The three projects presented in the current thesis involve both X-ray crystallography and molecular modeling techniques:

Aiming at the better understanding of the functional principles of the anticancer antibiotic Actinomycin D (AMD), the complex of its derivative 7-Aminoactinomycin D with non-complementary DNA was investigated via X-ray crystallography. Its interaction with double stranded DNA had been investigated thoroughly before, but no X-ray structure of a complex with single stranded DNA was known so far. The importance of such a structure lies not only in the fact that it would contribute insights into the anticancer action of AMD. Recent studies have revealed a possibility of the use of AMD against HIV-1 diseases through its interaction with ssDNA.

In the second project the structure of an alternating d,l-peptide has been elucidated. Most peptides of this type show a common folding motif: they are often found to be β -helices. The most prominent example is Gramicidin A, which acts as an ion channel antibiotic for specific monovalent cations. The peptide examined in this thesis is H-(L-Tyr-D-Tyr)₄-Lys-OH and was found to show the same structural motif. Its ability to form ion conductors in membranes is being further investigated.

The last project focuses on the theoretical investigation of human estrogen receptor hER α and the development of possible mutation sites in its binding pocket. With the help of molecular modeling techniques a new regulatory system for plant gene expression is developed using a mutant of the hER α ligand binding domain as a chemical switch that can be induced by organic ligands.

Focusing on three diverse, yet methodologically and applicationably related subjects, the work presented in this thesis exemplifies the wide range and power of structural biology, in particular X-ray crystallography, in the field of drug design.

THEORETICAL BACKGROUND

2.1 CRYSTALLOGRAPHIC METHODS

2.1.1 X-RAY DIFFRACTION BY CRYSTALS

X-ray scattering by crystals is a result of the interaction of electromagnetic waves with the electrons of the crystal atoms. If an electromagnetic wave hits one atom, its electrons will start to oscillate either with the same wavelength as the incident radiation (elastic scattering) or with a wavelength slightly higher (Compton radiation). In the first case, the electrons emit radiation that is responsible for the interference effects in diffraction. The scattering factor f_a of an atom depends on the electron density, the reflecting angle θ and the X-ray wavelength λ and can be described by the following equation:

$$f_a = \int_0^{\infty} U_a \frac{\sin\left(2\pi r \frac{2 \sin \theta}{\lambda}\right)}{2\pi r \frac{2 \sin \theta}{\lambda}} dr, \text{ where } U_a = 4\pi r^2 \rho_a(r).$$

By taking the temperature-dependent vibration of atoms into account a correction is introduced to f_a :

$$f'_a = f_a \exp\left(-B \frac{\sin^2 \theta}{\lambda^2}\right), \text{ where } B = 8\pi^2 u^2.$$

B is the atomic temperature factor, while u^2 is the squared displacement of the atom. Due to the limited resolution and therefore limited number of data it is common in protein crystallography to work with isotropic temperature factors for individual atoms.

When X-rays hit a crystal, the beam is scattered by a large number of electrons. The resulting waves interfere with each other. In certain orientations of the lattice planes, when all vectors point to one direction and the phase shift is $n\lambda$ positive interference takes place resulting in reflections, which constitute the diffraction pattern of the crystal. This is best described using *Bragg's* conception (Fig. 2.1).

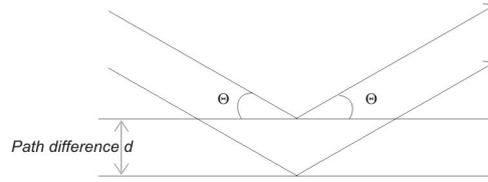


Fig. 2.1 The two lattice planes of the crystal are separated by the distance d . The X-rays are reflected by the planes with an angle θ .

A crystal can be described by a series of parallel diffracting planes separated by a distance d . These lattice planes are defined by the *Miller* indices h , k and l . The incident and reflected beam make an equal angle θ with the plane. Maximum positive interference of the waves is achieved when *Bragg's* law is fulfilled:

$$2d \sin \theta = n\lambda .$$

Due to the interference of the waves, the effective number of electrons in the scattered wave is reduced. The sum of the scattering by the atoms of the unit cell is then called *structure factor* F . For n atoms

$$F_{hkl} = \sum_{i=1}^n f_i' \{ \cos[2\pi(hx_i + ky_i + lz_i)] + i \sin[2\pi(hx_i + ky_i + lz_i)] \}$$

or

$$F_{hkl} = \sum_{i=1}^n f_i' \exp[2\pi i(hx_i + ky_i + lz_i)]$$

or

$$F_{hkl} = |F_{hkl}| \exp(i\varphi_{hkl}).$$

F is a complex number characterized by its amplitude $|F_{hkl}|$, whose square $|F_{hkl}|^2$ corresponds to the intensity of the reflection hkl , and its phase φ_{hkl} , which depends on the positions x_i , y_i , z_i of the atoms. Instead of summing over all atoms, one can integrate over all electrons in the unit cell:

$$F_{hkl} = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz .$$

The structure factor F_{hkl} represents the Fourier transform of $\rho(xyz)$ and vice versa, so $\rho(xyz)$ can be written as

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)].$$

By knowledge of the intensities and the phases of the reflections in the diffraction pattern, it is possible to calculate an electron density map for the unit cell. In reality, only the

intensities of reflections are measured during an X-ray measurement, while the phase shifts are missing. This problem is known as the *crystallographic phase problem*. For small molecules with up to 100 atoms the phases can be obtained by using *Direct Methods* techniques. However, in the case of macromolecular crystals, which possess a large number of atoms, solving the phase problem is still one of the major tasks crystallographers are confronted with.

2.1.2 SOLUTION OF THE PHASE PROBLEM

Four ways of overcoming the phase problem are used in macromolecular crystallography:

- a. Isomorphous replacement (2.1.2.2)
- b. The use of anomalous signal of heavy atoms incorporated in the crystal (2.1.2.3 & 2.1.2.4)
- c. Molecular replacement methods (2.1.3) that are exploiting *Patterson* search techniques
- d. *Ab initio* methods, which are purely statistic methods making use of probabilistic relationships

2.1.2.1 HEAVY ATOM DERIVATIVES

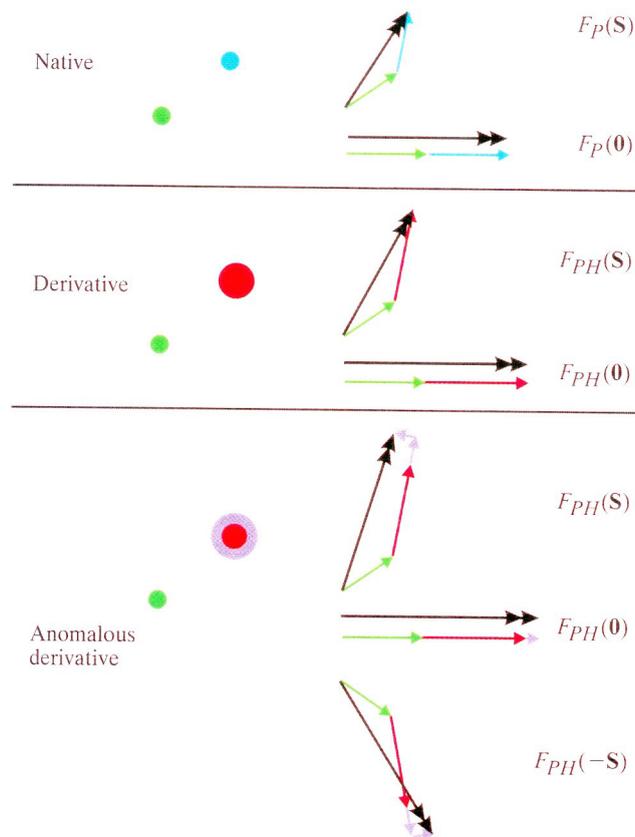


Fig. 2.2 Effects of the introduction of heavy atoms in a crystal. F_P describes the structure factor of the native structure, F_{PH} is the structure factor of the heavy atom derivative.

The introduction of heavy atoms in the crystal lattice has different effects on the structure factor F . In the example in Fig. 2.2 a two atom structure is examined. In the native structure the structure factor F_P is composed of the two contributions of the green and the blue atom. These atoms have the same magnitude but a different phase shift. By replacing the blue atom with a heavier red one its magnitude becomes greater, however the phase shift remains the same. This causes a change in the magnitude and the phase of the overall structure factor F_{PH} . The introduction of an anomalous scatterer results in a change of the magnitude and the phase of the structure factor F_H of the second atom, yielding different amplitudes and phases for $F_{PH}(S)$ and $F_{PH}(-S)$.

2.1.2.2 ISOMORPHOUS REPLACEMENT

Isomorphous replacement methods require the X-ray diffraction patterns of the native as well as the derivative protein.

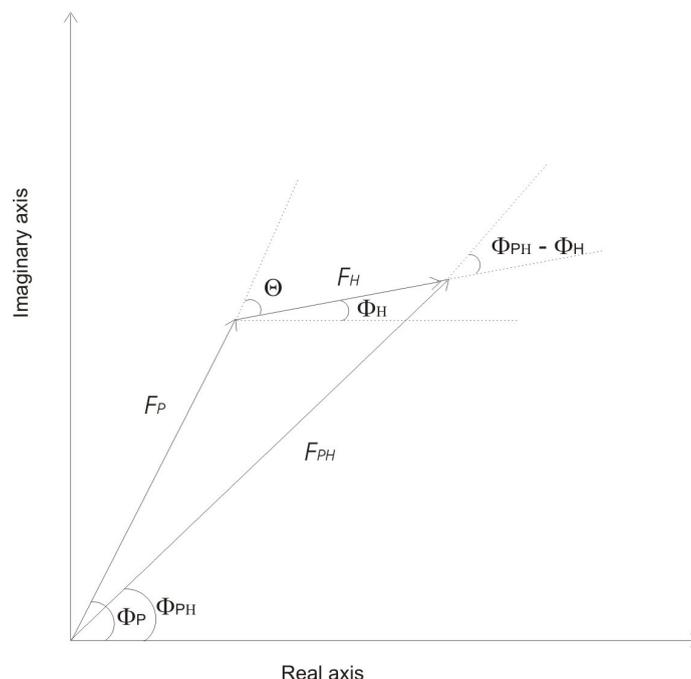


Fig. 2.3 Vector diagram showing the relationship among the structure factor of the native protein F_P , of the heavy atom F_H and of the heavy atom derivative F_{PH} .

If F_P is the structure factor of a native protein with the amplitude F_P and the phase Φ_P and the structure factor F_{PH} , its amplitude $|F_{PH}|$ and phase Φ_{PH} the corresponding quantities of the derivative structure, then:

$$F_{PH} = F_P + F_H,$$

where F_H is the structure factor contribution of the heavy atom derivative. The only parameters that can be measured in an X-ray experiment are the intensities $|F_P|$ and $|F_{PH}|$ and the aim is to derive the value of Φ_P . The factor F_H cannot be measured, but it can be calculated if the atomic coordinates and the thermal parameters of the heavy atoms are known.

The determination of the heavy atom coordinates in small molecule structures is done by calculating the *Patterson* map. The *Patterson* function is a Fourier summation with intensities as coefficients:

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)].$$

where u , v and w are relative coordinates to the unit cell. The *Patterson* function can be calculated by using only the measured intensities of the experiment and without any knowledge of the structure. For every interatomic vector (uvw) the *Patterson* map will contain a peak. As every atom makes an interatomic vector to itself, the origin of the map will dominate. However, it can be removed through subtraction of the average intensity $I(hkl)$ before Fourier transformation. Furthermore, the *Patterson* map is centrosymmetric and especially for bigger structures extremely crowded. Nevertheless, for small structures containing heavy atoms, after removal of the origin the map will be dominated by the peaks of the heavy atom interatomic vectors, which makes it easier to derive their positions.

In macromolecular heavy atom derivatives the ratio of the number of electrons of the heavy atoms to those of the protein is so low, that the *Patterson* map of the derivative does not provide any useful information. Alternatively, the *isomorphous difference-Patterson* map is calculated using coefficients $(|F_{PH}| - |F_P|)^2$.

As can be derived from Fig. 2.3:

$$F_H^2 = F_P^2 + F_{PH}^2 - 2F_{PH}F_P \cos(\varphi_P - \varphi_{PH}).$$

If the phases of protein and derivative structure are very similar, then the cosine term is close to 1, so following approximation can be made:

$$F_H^2 \approx F_P^2 + F_{PH}^2 - 2F_{PH}F_P = (F_{PH} - F_P)^2.$$

When the positions of the heavy atoms are known, F_H can be calculated. From trigonometric considerations it is clear that

$$F_{PH}^2 = F_P^2 + F_H^2 + 2F_PF_H \cos\theta, \text{ where } \theta = \varphi_P - \varphi_H,$$

so φ_P can be derived by

$$\varphi_P = \varphi_H + \cos^{-1} \left[\frac{(F_{PH}^2 - F_P^2 - F_H^2)}{2F_PF_H} \right].$$

The cosine term in this equation introduces a phase ambiguity and two solutions can be calculated. The Harker construction in Fig. 2.4 illustrates this. A circle with radius $|F_P|$ is intersecting with a second circle with the origin at the end of the vector $-|F_H|$ and with a radius of $|F_{PH}|$. The two intersection points represent the two possible solutions for φ_P . To solve this problem the multiple isomorphous replacement technique is applied, which makes use of several different heavy atom derivatives.

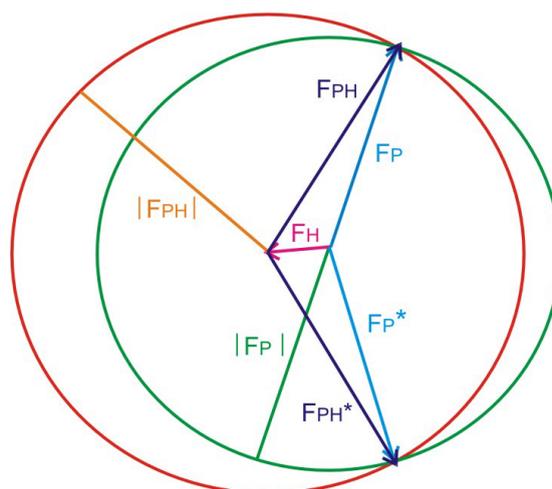


Fig. 2.4 Harker construction of the SIR method, showing the two possible values for the phase angle of the native structure factor F_p .

2.1.2.3 MAD

The electrons of an atom are bound to the nucleus, so they oscillate with a characteristic frequency. If the frequency of the incident X-ray beam is near this natural frequency resonance will take place. The scattering is then called anomalous and the scattering factor is replaced by a complex number and can be expressed as follows:

$$f = f_a + \Delta f' + if'' = f' + if''.$$

$\Delta f'$ and f'' are the real and imaginary dispersion corrections. The factor f' affects the scattering factor amplitude by reducing it, while the imaginary part f'' affects both phase and amplitude.

Anomalous scattering is in principle present for all atomic types at all X-ray energies. However, the magnitudes of f' and f'' are negligible if not near the resonant frequencies of the atom. This includes all “light” atoms (H, C, N, O) of biological macromolecules at all X-ray wavelengths commonly used in crystallographic experiments. Big effects have been measured for heavier atoms like rare-earth elements or halides (bromide, iodide). The plots of f' and f'' for the example of bromine are shown in Fig. 2.5.

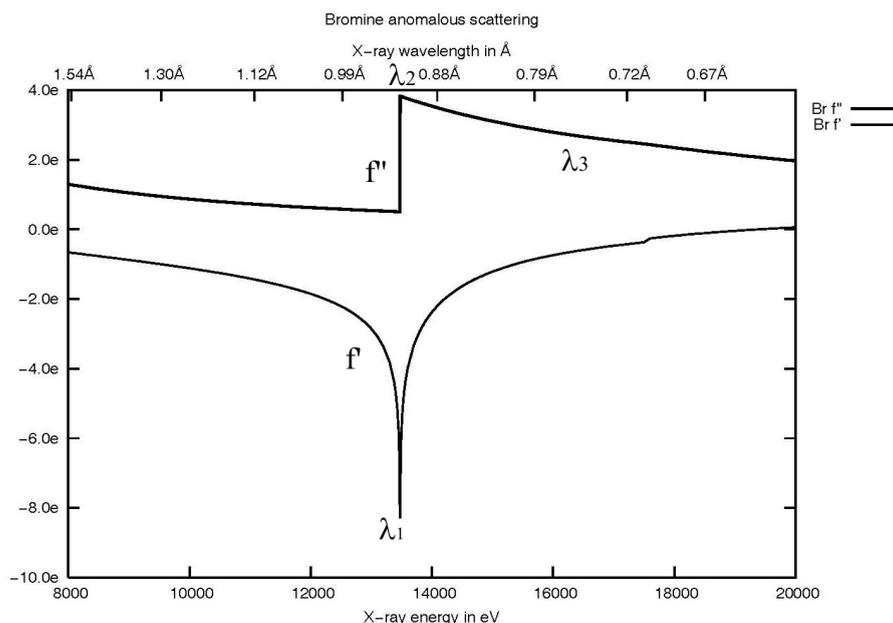


Fig. 2.5 The anomalous scattering curves (f' and f'') for bromine in the X-ray energy region around the absorption edge.

The position of the absorption edge of f' (λ_1 , edge) is the inflection point of f'' , while E_{peak} is the energy of peak absorption just above the edge (λ_2 , peak). In a MAD experiment, data sets are collected at λ_1 and λ_2 and often at a third wavelength λ_3 in the high-energy remote region.

Whenever there is anomalous scattering *Friedel's Law*:

$$|F_{hkl}| = |F_{-h-k-l}| \text{ and } \varphi_{hkl} = \varphi_{-h-k-l}$$

is not valid any more. The structure factor vectors in this case are illustrated in Fig. 2.6.

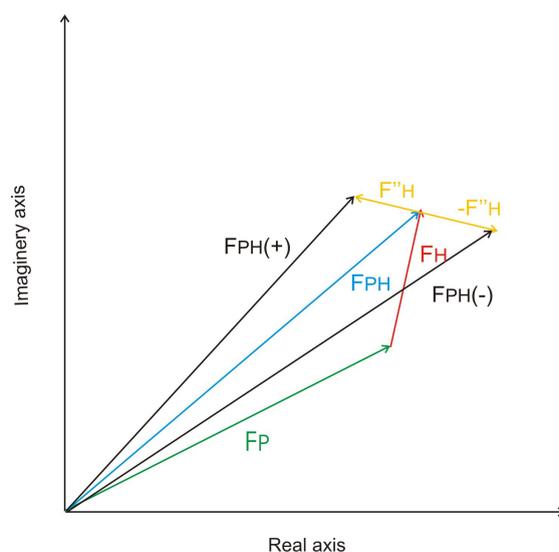


Fig. 2.6 Vector diagram showing $F_{PH(+)}$ and $F_{PH(-)}$, which have different magnitudes due to the anomalous contribution of the heavy atom.

The structure factors measured in a MAD experiment are the so called Bijvoet pairs $F_{PH}(+)$ and $F_{PH}(-)$, which result by addition of the anomalous correction factors F''_H and $-F''_H$ to F_{PH} respectively. The difference between the measured intensities of $F_{PH}(+)$ and $F_{PH}(-)$ is called a Bijvoet difference

$$\Delta|F| = |F_{PH}(+)| - |F_{PH}(-)|.$$

The *Patterson* map calculated with coefficients ΔF^2 contains only peaks corresponding to interatomic vectors between pairs of anomalous scatterers. With this *Bijvoet difference Patterson* the location of the anomalous heavy atoms can be found and used to estimate first phases.

The Karle/Hendrickson equations (Karle, 1980 and Hendrickson, 1985) express the magnitude of the measured intensities in terms of its components:

$$|F_{PH}(\pm)|^2 = |F_P|^2 + \frac{f''^2 + f'^2}{f_a^2} |F_A|^2 + 2 \frac{f'}{f_a} |F_P| |F_A| \cos \alpha \pm 2 \frac{f''}{f_a} |F_P| |F_A| \sin \alpha,$$

where $\alpha = \varphi_P - \varphi_A$, and F_A is the heavy atom structure factor. This equation has three unknown quantities $|F_A|$, $|F_P|$ and α . In most MAD experiments data sets are measured at three wavelengths λ_1 (inflection), λ_2 (peak) and λ_3 (high energy remote), so there results a system of six equations with three unknown variables.

2.1.2.4 SAD (Dauter et al, 2002, Dodson, 2003)

If only one set of Bijvoet pairs is available, it is not possible to estimate the protein phases exactly. This is the case in a SAD experiment performed at one wavelength. As in the SIR case there is a twofold phase ambiguity problem. The case of one kind of anomalous scatterers is illustrated in Fig. 2.7.

The two possible values for ϕ_{PH} are symmetrically oriented around $\phi_A - 90^\circ$. An equation can be deduced for the Bijvoet difference:

$$F_{PH}^2(+)-F_{PH}^2(-)=4F_{PH}F_A''\sin(\varphi_{PH}-\varphi_A).$$

For a small contribution of the anomalous atoms to the total diffracting power of the crystal, $F_A \ll F_{PH}$, and

$$\frac{|F_{PH}(+)|+|F_{PH}(-)|}{2} \approx F_{PH}$$

and

$$\Delta F_{PH}(\pm) \approx 2F_A'' \sin(\varphi_{PH} - \varphi_A).$$

Since

$$\sin(\varphi_{PH} - \varphi_A) = \sin(180^\circ - \varphi_{PH} + \varphi_A)$$

it is

$$\varphi_{PH} = \varphi_A + 90^\circ \pm \theta, \text{ with } \theta = \cos^{-1}(\Delta F_{PH}(\pm) / 2F_A'').$$

The probability of the phase distribution can be expressed by following equation:

$$P_{anom}(\varphi) = N \exp\left\{-\left[\Delta F_{PH}(\pm) + 2F_A'' \sin(\varphi_{PH} - \varphi_A)\right]^2 / 2E^2\right\}.$$

with N = normalizing factor and E = standard error estimation. The sine allows two possible values for $P_{anom}(\varphi)$, however there is a slightly higher probability for the phase φ_{PH} to have a value close to φ_A . Therefore Ramachandran and Raman (1956) suggested for φ_{PH} the choice of the value that is closest to φ_A for initial phasing.

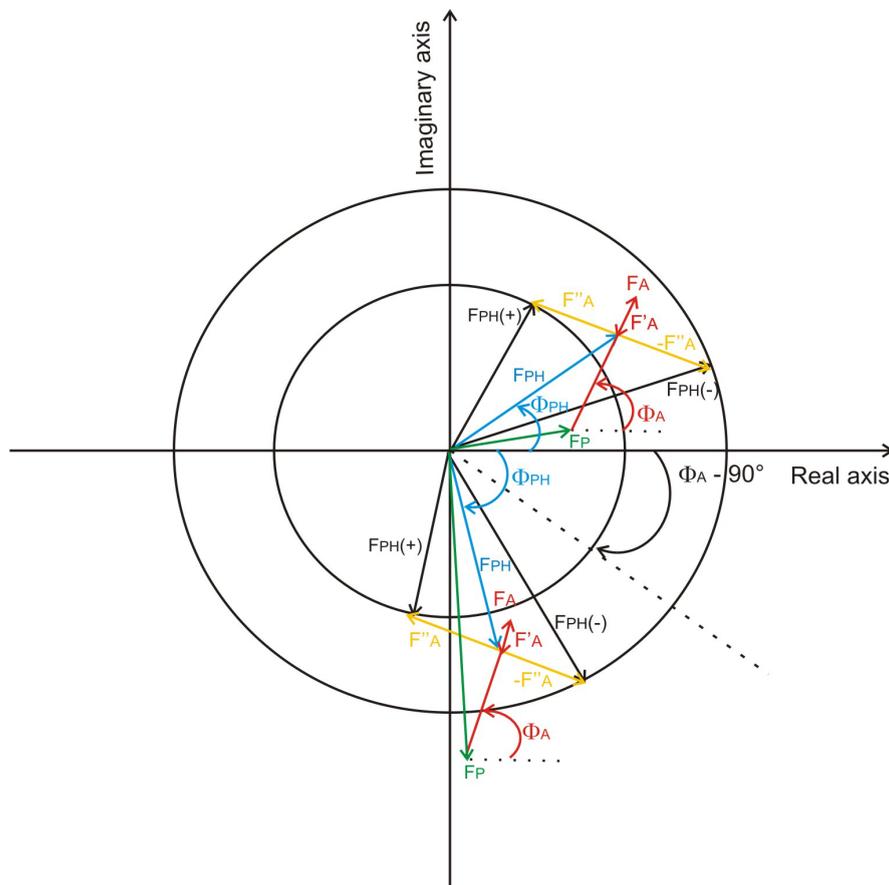


Fig. 2.7 The Argand diagram showing various contributions to the scattering factors. The contribution of normally scattering atoms is shown in green (F_P), the contribution of the anomalous scatterers F_A and F'_A is shown in red and that of F''_A in yellow, the total contribution F_{PH} is shown in blue and the amplitudes of both Friedel mates ($F_{PH(+)}$ and $F_{PH(-)}$) is shown in black.

2.1.2.5 SUBSTRUCTURE SOLUTION WITH SHELXD

After the determination of the heavy atom structure factors, SHELXD (Sheldrick & Usón, 2001) can be used to find the heavy atom coordinates. A phasing algorithm is applied for this purpose, which has been partly taken over from the *Shake-and-Bake* method introduced by Miller et al. (1994) and is implemented in the program SnB (Fig. 2.8).

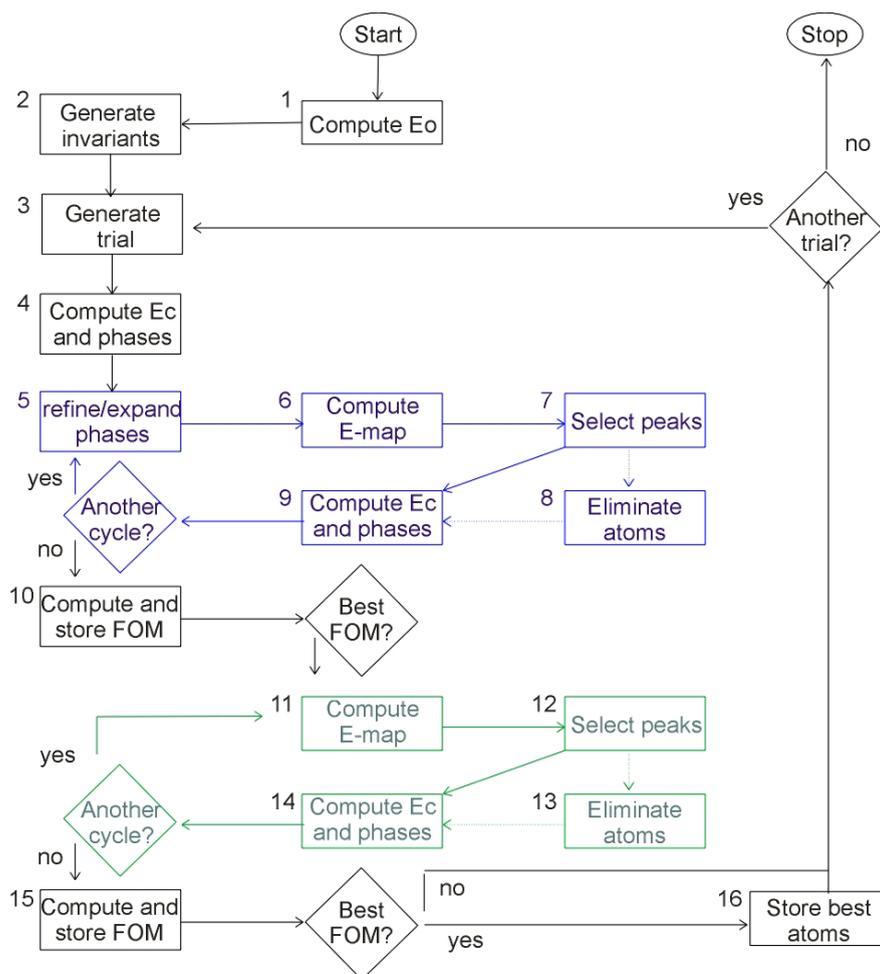


Fig. 2.8 Flowchart for the *Shake-and-Bake* procedure implemented in the programs SHELXD and SnB.

The *Shake-and-Bake* algorithm contains two refinement loops for the randomly generated initial phases. Both loops (Fig. 2.8 blue cycles 5-9, green cycles 11-14) are performed in reciprocal space and in real space. In the reciprocal space (*shaking*) either Karle-type tangent expansion or parameter-shift refinement is performed. If the FOM (correlation coefficient) after steps 5-9 is within a specified threshold (1-5%), the real space loop (*baking*) is entered by calculating an electron density map with a *Fast-Fourier-Transform* algorithm. In this map, the strongest peaks are picked and structure factors are re-calculated

from their positions. This procedure of dual-space steps is continued and validated by using a reciprocal space correlation coefficient.

2.1.3 MOLECULAR REPLACEMENT

When a good model for a reasonably large fraction of the structure in a crystal is available, phases can be obtained by using the method of Molecular Replacement (MR). The higher the sequence identity between the model and the unknown protein, the higher the success rate for MR. As a rule of thumb, the molecule used should be fairly complete and share at least 40% sequence identity with the unknown structure. In the last few decades the number of solved structures in the protein data bank has increased in a nearly exponential manner, so MR will be useful for a larger fraction of new structures.

The main task that has to be accomplished is the correct placement of the known model in the target unit cell. Six variables are used for this transformation: three rotational and three translational. During the rotation procedure, the orientation of the known and unknown structure is determined with respect to each other. In the translation step, the correctly oriented molecule is superimposed onto the other.

Traditional molecular replacement programs are based on the properties of the *Patterson* function. The *Patterson* map contains peaks that correspond to vectors between atoms of the same molecule (*self-Patterson vectors*) and to vectors between atoms of different molecules (*cross-Patterson vectors*). In the first case, the peaks are near the origin of the map, as the vectors are relatively short. This inner region of the *Patterson* map is equal for the symmetry equivalent molecules of the same crystal and for the same molecules in different crystal structures or similar for homologous molecules. The only difference is a rotational correction. As regards the *cross vectors*, they provide information about the correct position of the model in the unit cell.

In most MR programs, the orientation is determined first followed by the translation search. In the rotational step the *Patterson* map is superimposed onto rotated versions of itself. If the function is superimposed on a correctly rotated version, the overlap between the two maps will be maximal, providing the right orientation of the model. Especially if the *Patterson* function is restricted to a region Ω of volume V centered at the origin and having a dimension in the order of the isolated molecule, the *self-Patterson* terms will dominate the *crossed-Patterson* ones. The classical rotation function is commonly defined as a product function:

$$\mathfrak{R}(R) = \frac{1}{V} \int_{\Omega} P_t(r) P_s(R^{-1}r) d^3r,$$

where P_t is the target *Patterson* function, P_s the rotated search *Patterson* function and R the rotation.

There are different ways of plotting the magnitude of the rotation function in three-dimensional space. Usually two alternative systems are employed, the Eulerian coordinate system (Fig. 2.9 a) and the polar coordinate system (Fig. 2.9 b).

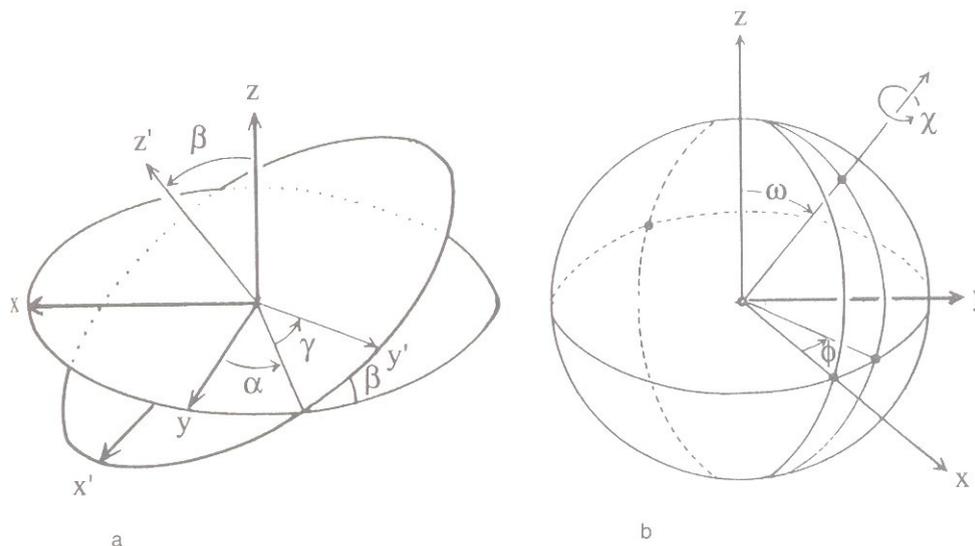


Fig. 2.9 a) Eulerian angles α , β and γ , b) Polar angles ϕ , χ and ω .

As the *Patterson* function is mainly determined by the strong reflections, it is important for the rotational step that all strong reflections are used. Furthermore, the best resolution range is between 4 and 8 Å. In this region the low resolution data are excluded as they are rather insensitive to rotation and the high resolution data are removed as they are very sensitive for the model.

As soon as the correct orientation is found, the translational search can begin. In this step, the correctly oriented molecule is moved through the asymmetric unit. For every position structure factors are calculated (F_{calc}) and compared with the observed ones either by calculating an R -factor

$$R = \frac{\sum_{hkl} \|F(obs) - k|F(calc)\|}{\sum_{hkl} |F(obs)|}$$

or a correlation coefficient

$$CC = \frac{\sum_{hkl} \left(|F(obs)|^2 - \overline{|F(obs)|^2} \right) \times \left(|F(calc)|^2 - \overline{|F(calc)|^2} \right)}{\sqrt{\sum_{hkl} \left(|F(obs)|^2 - \overline{|F(obs)|^2} \right)^2 \sum_{hkl} \left(|F(calc)|^2 - \overline{|F(calc)|^2} \right)^2}}.$$

2.1.4 DENSITY MODIFICATION

Unless the data have a very good quality and a high resolution, the initial phases obtained after solving a structure are quite crude and need to be optimized. This is done with density modification methods, which include *solvent flattening*, *histogram matching* and *NCS averaging*.

Solvent flattening: In macromolecular crystals there is often a high percentage of solvent, which can vary from 30-80%. The solvent is found in the crystal lattice mostly in the form of solvent channels between ordered protein molecules. Its nature is highly dynamic resulting in a low constant value for its time-averaged electron density. Therefore it is possible to flatten the solvent region by removing its noise peaks and setting its electron density to a low value. In the Wang method (1985), a three-dimensional grid is superimposed on the unit cell to determine the solvent mask. The electron density for each grid point j in a sphere of radius R containing grid points i is calculated with the following equation

$$\rho'_j = K \sum_i^R w_i \rho_i,$$

where K is an arbitrary constant. For $\rho_i < 0$, $w_i = 0$ and for $\rho_i > 0$, $w_i = 1 - (r_{ij}/R)$.

Histogram matching: Zhang and Main (2001) systematically examined the electron-density histograms of several proteins and found that the ideal density histogram is independent of structural conformation, but is regulated by the resolution, the overall temperature factor B_{over} and the phase error. The last point implies that the phases can be improved by histogram modification, when the resolution and B_{over} are known. The prediction of the ideal histogram for protein structures is based upon the similarities observed. These include the similar atomic composition and the conserved bond formation. Histogram matching can be used for phase extension.

NCS averaging: It is fairly common for protein crystals to contain more than one molecule in the asymmetric unit related by one or more symmetry operators. These operators do not apply on the whole crystal and are therefore called non-crystallographic symmetry (NCS).

NCS operations can often be identified in the *self-Patterson* rotation function. The molecules related by NCS show an almost equal electron density, apart from the regions of contact with neighboring copies. This information can be used as a constraint for the protein structure factors. As a first step the NCS operators are determined. The electron density of the corresponding molecules is then chosen, mostly slightly smaller than the experimental one, in order to exclude neighbor contact regions. This density is averaged and after solvent flattening the asymmetric unit is reconstituted. By back-Fourier-transform new phases are calculated, which should lead to a better map, so the NCS operators can be further refined. This procedure can be repeated for several cycles.

SHELXE: In the *sphere of influence* algorithm incorporated in SHELXE the variance V of density on a spherical surface of radius 2.42 Å is calculated for each pixel in the map, and the pixels with the highest varieties are assigned to the protein region, while the others belong to the solvent. In that way some chemical information is added with 2.42 Å being the typical 1,3-distance in proteins and DNA. The variance of V over all pixels is defined as the *contrast*. The *connectivity* is the fractions of adjacent pixels that are both either in the solvent or both in protein regions. Last a *pseudo-free correlation coefficient* is calculated leaving out 10% of reflections at random and doing one cycle of density modification.

2.1.5 STRUCTURE REFINEMENT

The initial interpretation of the electron density map of macromolecular structures at medium resolution derived after structure solution even when improved with density modification is in general not easy. In the case of proteins, programs are available (ARP/wARP) that can automatically trace the polypeptide chain if the sequence is known. However, for other macromolecules such as antibiotics or DNA/RNA structures the tracing has to be done manually, which can be very time-consuming.

Once a model has been built into the map, it needs to be adjusted so that a closer agreement between the calculated and the experimental structure factors is obtained. This agreement is given by the R -factor

$$R = \frac{\sum_{hkl} \|F(obs) - k|F(calc)\|}{\sum_{hkl} |F(obs)|}.$$

A minimization of the R -factor is achieved by optimizing the model.

The refinement technique used in this thesis is based on the principle of *least-squares*. The observations have fixed values and the parameters are varied such that the calculated values approach the observations as closely as possible. The limits in the macromolecular crystal diffraction result in a rather poor observation to parameter ratio. It is therefore not sufficient just to refine the positions and *B*-factors of the atoms. The solution is either to add “observations” in the form of restraints or to reduce the number of parameters by using constraints.

Refinement by *least-squares* is performed iteratively and a great number of refinement cycles have to be carried out until the changes in the parameters have converged. The function that is actually minimized is

$$Q = \sum_{hkl} w(hkl) \left(|F_{obs}(hkl)| - |F_{calc}(hkl)| \right)^2 .$$

All crystallographically independent reflections are included in the summation and the restraints are taken into account by the weighting factor w . To reach the minimum for Q , its differentials with respect to all atomic parameters u_j are set to zero

$$\sum_{hkl} w(hkl) \left(|F_{obs}(hkl)| - |F_{calc}(hkl)| \right) \frac{\partial |F_{calc}(hkl)|}{\partial u_j} = 0 .$$

F_{calc} is expressed in a Taylor expansion in which the starting values u_s change by a small amount ε and the higher order terms are neglected. These ε values must be found and applied to the variables u_j . There results a system of normal equations that can be written in the matrix form

$$\begin{bmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} & \dots \\ \alpha_{12} & \alpha_{22} & \alpha_{32} & \dots \\ \alpha_{13} & \alpha_{23} & \alpha_{33} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \times \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \dots \end{bmatrix}$$

or

$$[A] \times [\varepsilon] = [b],$$

where $[A]$ is called the normal matrix with the elements

$$\sum_{hkl} w(hkl) \left[\frac{\partial |F_{calc}(hkl; u)|}{\partial u_i} \right]_{u_s} \times \left[\frac{\partial |F_{calc}(hkl; u)|}{\partial u_j} \right]_{u_s} ,$$

and $[b]$ is the known gradient vector containing the elements

$$\sum_{hkl} w(hkl) \left(|F_{obs}(hkl)| - |F_{calc}(hkl; u_s)| \right) \times \left[\frac{\partial |F_{calc}(hkl; u)|}{\partial u_j} \right]_{u_s} .$$

To reduce the computational time, a simplification is made that all off-diagonal elements of [A] are zero. Nevertheless, in almost all cases there is geometric or energy information available besides the X-ray data. This is incorporated in the refinement by adding extra terms to the matrices [A] and [b]. In *conjugate gradient* algorithm (Tronrud, 1992) initial values are estimated for the parameter shifts $[\epsilon_0]$ and a residual vector matrix $[r_0] = [b] - [A] \times [\epsilon_0]$ is calculated.

To avoid over-interpretation of the data by a too elaborate model (*model bias*) the method of *cross validation* (Brünger, 1992) was introduced. For this method, the diffraction data is divided into two sets: a large *working* set (mostly comprising 95% of the data) and a small *free* data set (with the remaining 5%). The latter is not used during the refinement process, but is a good indicator of model quality.

2.2 MOLECULAR MODELING

The determination of a structure via X-ray crystallography is one of the most reliable methods to obtain a three-dimensional structure. However, especially for proteins it is quite time-consuming, including protein expression and purification, crystallization, data collection, structure solution and refinement. Each one of these steps can require several months to years. This is a handicap that affects a lot of relevant fields. In particular it can represent a hindrance in the process of drug design, where the interactions between target and drug are investigated. The investigation of the specific interaction of a number of ligands with biologically relevant macromolecules would demand years. In order to avoid this time problem, computer models of proteins or protein complexes have been used in the last ten to twenty years, by using a technique called molecular modeling. It represents molecular structures numerically and simulates their behavior using equations of quantum and classical physics. Geometries, energies, electronic properties, spectroscopic properties and bulk properties can be generated and presented. In this thesis the molecular modeling consists of two steps: a) docking of organic ligands in a binding site and b) energy minimization of the resulting complexes.

2.2.1 DOCKING WITH FLEXX

FlexX (Rarey et al., 1996) is a computer program for predicting protein-ligand interactions. For a given protein and a ligand FlexX predicts the geometry of the complex as well as the

free energy. The algorithm used is based on the work of Böhm (Böhm, 1992a, 1994) and Klebe & Mietzner (1994).

2.2.1.1 THEORETICAL ASPECTS

The ligand input structure should already be reasonably minimized as FlexX keeps its bond lengths and angles. The conformational flexibility is modeled by variation of the torsion angles. A set of up to 12 preferred torsion angles is assigned to each acyclic single bond. The angles selected have a high population in the occurrence histogram of the Cambridge Structural Database (CSD) and should therefore lead to the generation of low energy conformations. As regards the protein-ligand interactions, FlexX uses mainly hydrogen bonds, but also hydrophobic and ionic (metal) interactions. The geometrical restriction is described by four types of interaction surfaces (Fig. 2.10). Each interacting group of the molecule is assigned an interaction type and an interaction geometry.

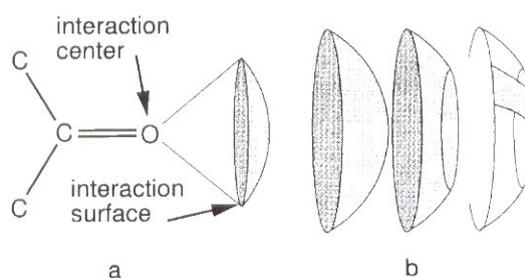


Fig. 2.10 Interaction geometries a) The carbonyl group is defined as the interaction center with its corresponding interaction surface, b) Three of the four different interaction surfaces types: cones, capped cones and spherical rectangles.

An interaction between two groups *A* and *B* is formed if

- (a) The interaction types of *A* and *B* are compatible
- (b) The interaction center of *A* lies approximately on the interaction surface of *B* and vice versa (Fig. 2.11).

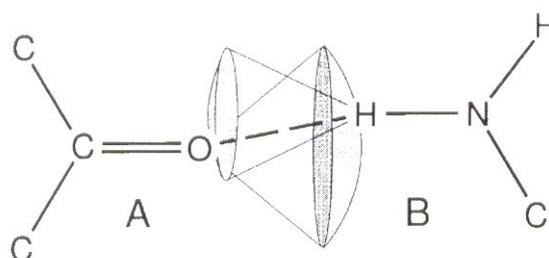


Fig. 2.11 The interacting centers O and H form a hydrogen bond and fall mutually on the surrounding interaction surfaces.

After the docking of the ligands in the active site, the generated solutions are ranked using a scoring function that estimates the free binding energy ΔG of the protein-ligand complex:

$$\Delta G = \Delta G_0 + \Delta G_{rot} \times N_{rot} + \Delta G_{hb} \sum_{neutralhb} f(\Delta R, \Delta \alpha) + \Delta G_{io} \sum_{ionic} f(\Delta R, \Delta \alpha) + \Delta G_{aro} \sum_{aro} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} \sum_{lipo} f^*(\Delta R)$$

where N_{rot} is the number of free rotatable bonds and $f(\Delta R, \Delta \alpha)$ is a scaling function penalizing deviations from the ideal geometry. Hydrogen bonding, ionic, aromatic and lipophilic interactions are taken into account with the adjustable parameters ΔG_{hb} , ΔG_{io} , ΔG_{aro} and ΔG_{lipo} .

2.2.1.2 ALGORITHM

The docking in FlexX follows an incremental construction strategy, which consists of three steps:

- Base selection. The first phase of the docking algorithm is the selection of a connected part of the ligand, the base fragment.
- Base placement. The base fragment is positioned into the active site independently of the rest of the ligand.
- Complex construction. In this last step the ligand is constructed in an incremental way, starting with the different placements of the base fragment.

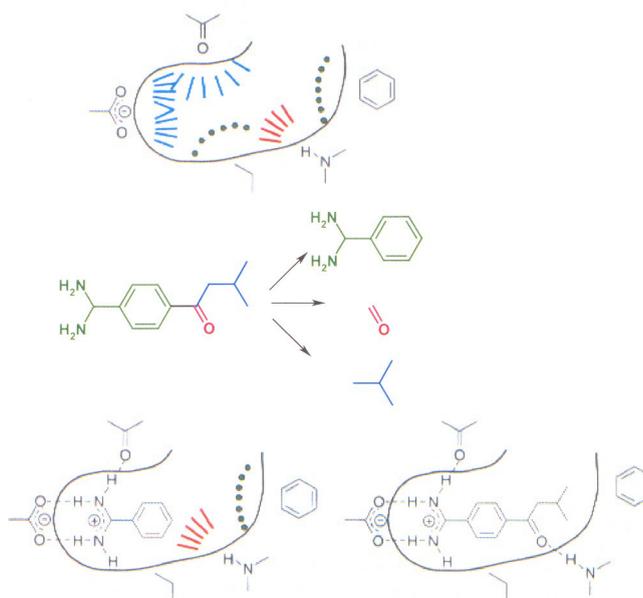


Fig. 2.12 The FlexX approach: Firstly, the interaction sites of the binding niche are calculated. As a second step, the ligand is defragmented and the fragments are docked and then linked. As a last step, the scoring function is calculated for the solutions.

Once the base fragment has been selected, the rest of the molecule is defragmented as well. The fragments are rather small, as the ligand is cut at each acyclic single bond. In the base placing step, the goal is to find positions of the base fragment in the active site so that a sufficient number of favorable interactions between the fragment and the protein can occur simultaneously. To achieve this, the ligand transformation into the binding pocket is defined by simply superimposing three interaction centers of the base fragment onto three interaction points of the receptor (Fig. 2.13).

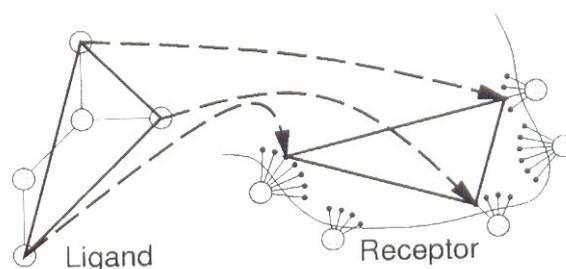


Fig. 2.13 Superimposition of the interaction centers of the ligand on the interaction points of the receptor.

The placement of the base fragment in the pocket is monitored by two filters: First it is checked if the interaction centers of the receptor coincide with the interaction surfaces of the ligand base fragment and then the base fragment is checked for overlap with the receptor. After a set of favorable placements for the base fragment has been computed, the incremental construction process for the ligand can begin. The remaining fragments are then added to the alternative placements of the base part. The strategy is called the *tree search problem*. On the first level of the tree are the first placements of the base fragment. Further fragments are linked to the first part, so branching takes place. The goal is to find the leaves which contain placements with favorable binding energies as estimated by the scoring function. It is not possible to carry out a complete search of the tree, as the number of degrees of freedom increases exponentially. Alternatively, the energy of partially placed ligands is used for filtering out unfavorable placements and keeping the others.

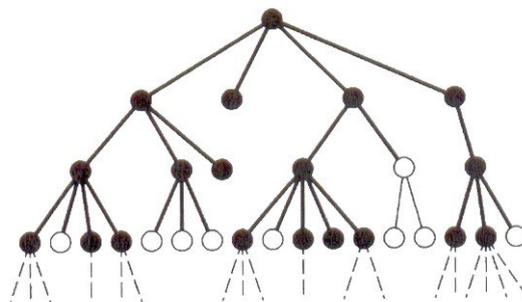


Fig. 2.14 Tree search process of FlexX.

After choosing a chemically sensible solution from the possibilities calculated by FlexX, the energy of the complex can be minimized with the program Discover 2.98.

2.2.2 ENERGY MINIMIZATION WITH DISCOVER

One possibility of the mathematical description of a molecule is the non-relativistic Schrödinger equation

$$H\Psi(R, r) = E\Psi(R, r),$$

where H is the Hamiltonian for the system, Ψ is the wave function and E is the energy. In the Born-Oppenheimer approximation the motion of the electrons can be decoupled from that of the nuclei, so two separate equations are obtained. The equation describing the electronic motion depends only parametrically on the positions of the nuclei

$$H\psi(r; R) = E\psi(r; R).$$

It defines an energy $E(R)$, which is a function only of the coordinates of the nuclei, called *potential energy surface*.

In the second equation the motion of the nuclei on the potential energy surface $E(R)$ is described

$$H\Phi(R) = E\Phi(R).$$

By solving the first equation a value is obtained for $E(R)$ that can be used in the second one. Nevertheless, the effort to solve the electron motion equation is extremely large so an empirical fit to the potential energy surface is used. This empirical fit is the *force field*. The force field defines the coordinates used and the parameters adjusted. The goal is to describe entire classes of molecules with reasonable accuracy. The force field used in this thesis is the CFF91 force field (Fig. 2.15). It employs a quartic polynomial for bond stretching (term 1) and angle bending (term 2) and a three-term Fourier expansion for torsions (term 3). An out-of-plane coordinate is incorporated in term 4. Seven cross terms follow (5-11), the Coulomb interaction between the atomic charges described in term 12 and term 13 that represents the van der Waals interactions.

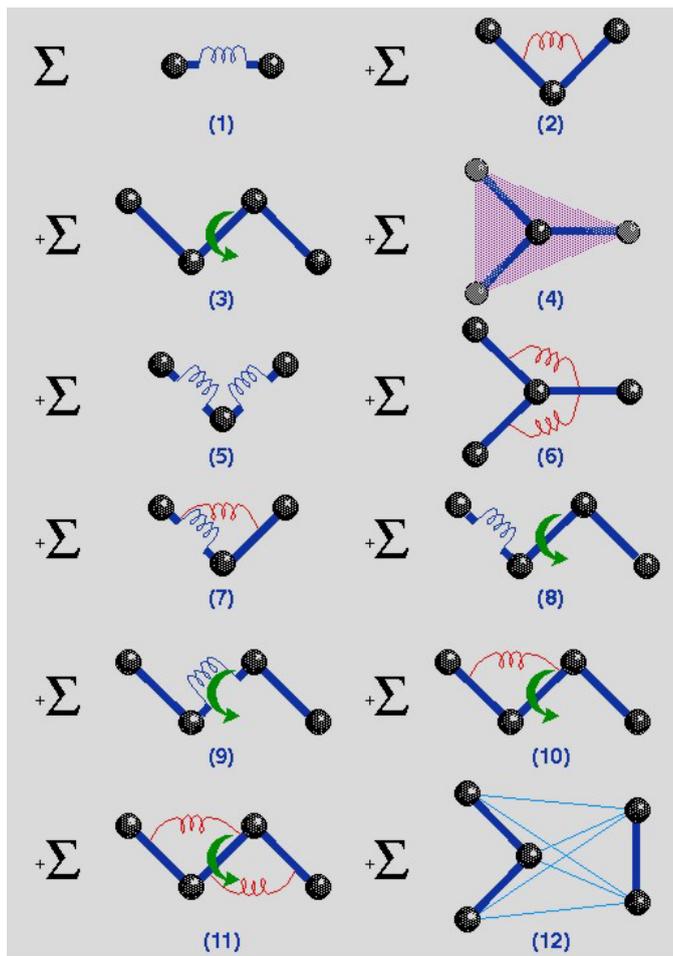


Fig. 2.15 Illustration of terms in force field CFF91.

The combination of the force field data with the coordinates of the molecule create the *energy expression* or *target function* for the molecule. For CFF91 it is

$$\begin{aligned}
 E_{pot} = & \sum_b [k_2(b-b_0)^2 + k_3(b-b_0)^3 + k_4(b-b_0)^4] \\
 & + \sum_\theta [H_2(\theta-\theta_0)^2 + H_3(\theta-\theta_0)^3 + H_4(\theta-\theta_0)^4] \\
 & + \sum_\varphi \{V_1[1-\cos(\varphi-\varphi_1^0)] + V_2[1-\cos(2\varphi-\varphi_2^0)] + V_3[1-\cos(3\varphi-\varphi_3^0)]\} \\
 & + \sum_\chi K_\chi \chi^2 + \sum_b \sum_{b'} F_{bb'}(b-b_0)(b'-b_0') + \sum_\theta \sum_{\theta'} F_{\theta\theta'}(\theta-\theta_0)(\theta'-\theta_0') \\
 & + \sum_b \sum_\theta F_{b\theta}(b-b_0)(\theta-\theta_0) + \sum_b \sum_\varphi (b-b_0)(V_1 \cos \varphi + V_2 \cos 2\varphi + V_3 \cos 3\varphi) \\
 & + \sum_{b'} \sum_\varphi (b'-b_0')(V_1 \cos \varphi + V_2 \cos 2\varphi + V_3 \cos 3\varphi) \\
 & + \sum_\theta \sum_\varphi (\theta-\theta_0)(V_1 \cos \varphi + V_2 \cos 2\varphi + V_3 \cos 3\varphi) \\
 & + \sum_\varphi \sum_\theta \sum_{\theta'} K_{\varphi\theta\theta'} \cos \varphi (\theta-\theta_0)(\theta'-\theta_0') + \sum_{i>j} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{i>j} \left(\frac{A_{ij}}{r_{ij}^9} - \frac{B_{ij}}{r_{ij}^6} \right)
 \end{aligned}$$

This energy expression is computationally tractable only for systems with a relatively small number of atoms, as the number of internal coordinates grows linearly with the size of the molecule. For macromolecules, a non-bond cutoff is introduced, which means that non-bond interactions are neglected for pairs of atoms separated by distances greater than a cutoff value.

By minimizing this energy stable conformations of a molecule can be identified. In a first step, the energy is calculated for a given conformation. Then the conformation is adjusted to lower the value of the target function. A minimum may be found after just one adjustment, but may require many thousands of iterations. This depends on the size of the molecule, the nature of the algorithm used and the form of the target system.

The algorithms used by *Discover* are *Steepest Descent* and *Conjugate Gradient*. The steepest descent minimization path is shown in Fig. 2.16.

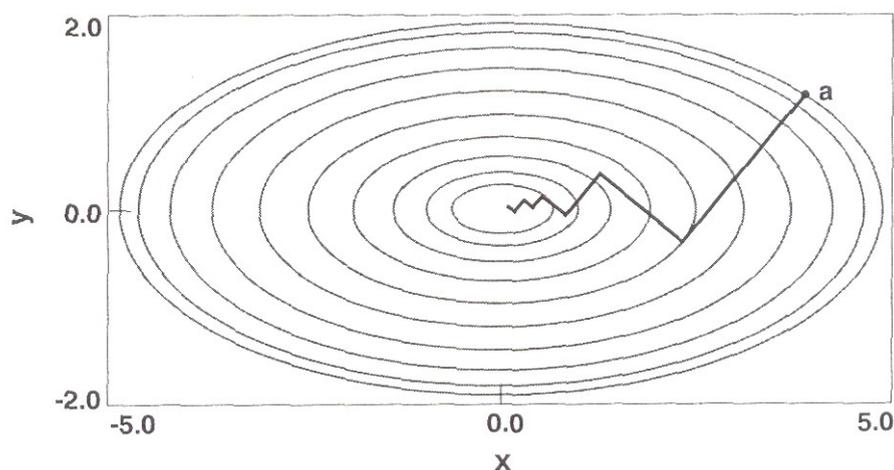


Fig. 2.16 Minimization following a steepest descent path for a quadratic function.

In this method a line search direction is followed along the direction of the local downhill gradient $-\nabla E(x_i, y_i)$. As soon as the minimum of that path is reached, the direction is changed and a new line search is started orthogonal to the previous one. Convergence for this technique is slow near the minimum because the gradient approaches zero, but the method is very robust. Steepest descents is often used when the gradients are large and the configurations are far from minimum.

To avoid the path oscillation of the steepest descents, the direction of the new iteration is controlled. This is incorporated in the method of *conjugate gradient*. The time per iteration for conjugate gradient functions is much longer than for steepest descents, but the convergence is far more efficient.

THE 7-AMINOACTINOMYCIN D COMPLEX WITH NON-COMPLEMENTARY DNA

3.1 INTRODUCTION

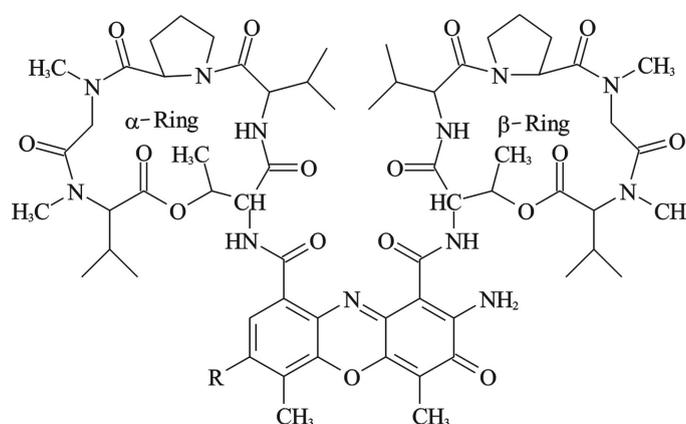


Fig. 3.1 R = H: Chemical structure of Actinomycin D
R = NH₂: Chemical structure of 7-Aminoactinomycin D.

The drug Actinomycin D (AMD) (Fig. 3.1) is a highly potent antitumor agent, which generates a wide variety of biochemical and pharmacological effects. It is commonly used in the treatment of rhabdomyosarcoma and Wilm's tumor in children as a component of the clinical VAC therapy (vincristine, actinomycin D, cyclophosphamide). The pharmacological function of AMD can be traced to its interactions with DNA, and in particular to its ability to inhibit the transcription of DNA to RNA.

Because of its interesting sequence specificity, the AMD/DNA complex has been well studied by thermodynamics, foot-printing, X-ray crystallography, NMR and spectroscopy. The structures of AMD complexes with DNA deposited in the Protein Data Bank are listed in Table 3.1.

PDB ID	Method	Resolution	Species involved	Corr. Author
173D	x-ray	3 Å	AMD-d(GAAGCTTC)	<i>Kamitori & Takusagawa, 1994</i>
1I3W	x-ray	1.7 Å	AMD-d(CGATCGATCG)	<i>Wang et al, 2001</i>
209D	x-ray	3 Å	N8-AMD-d(GAAGCTTC)	<i>Takusagawa et al, 1995</i>
2D55	x-ray	3 Å	AMD-d(GAAGCTTC)	<i>Takusagawa et al, 1995</i>
316D	x-ray	3 Å	F8-AMD-d(GAAGCTTC)	<i>Takusagawa et al, 1997</i>
1MNV	x-ray	2.6 Å	AMD-d(ATGCTGCAT)	<i>Wang & Hou, 2002</i>
1DSC	NMR		AMD-d(GAAGCTTC)	<i>Wang et al, 1996</i>
1DSD	NMR		AMD-d(GATGCTTC)	<i>Wang et al, 1996</i>
1FJA	NMR		AMD-d(AAGCGCTT)	<i>Chen & Patel, 1996</i>
1L1V	NMR		AMD-d(GTCACCGAC)	<i>Chen et al, 2002</i>

Table 3.1 Structures of Actinomycin D complexes in the Protein Data Bank.

These studies have revealed that 5'-GpC-3' sequence is the major binding site, with the phenoxazone ring of AMD intercalating into the 5'-GC/CG-5' step and the two cyclic pentapeptide lactones anchoring on both sides of the minor groove. The recognition of the binding site is achieved by hydrogen bonding between the threonines of the AMD and the guanines of the intercalation site. Furthermore, the guanine rings of G-C base pairs overlap dominantly on the phenoxazone ring whereas the cytosine rings show very little overlap. Additional stabilization of the complex results from hydrophobic interactions between the cyclic pentapeptides and the DNA minor groove surface atoms (Fig. 3.2).

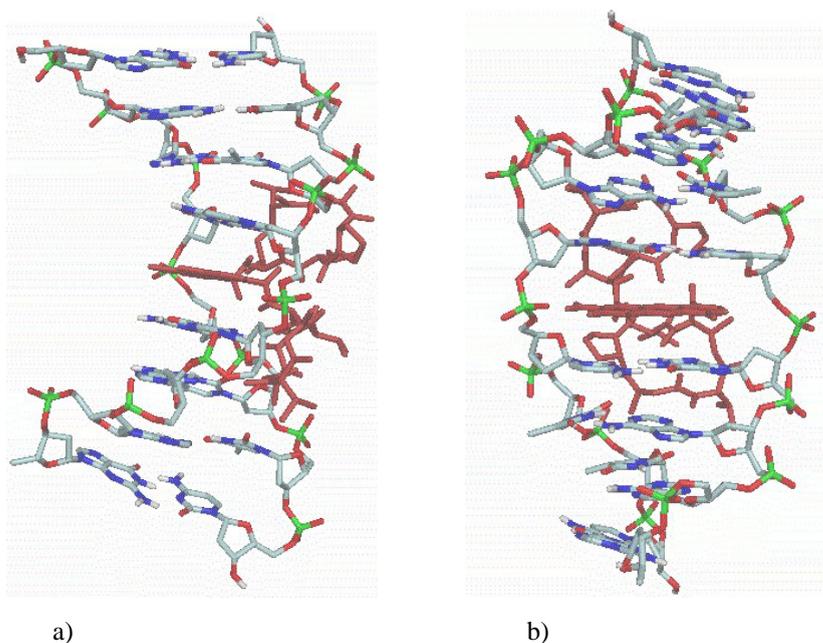


Fig. 3.2 Side (a) and front (b) view of the crystal structure of the 1:1 complex AMD:d(GAAGCTTC).

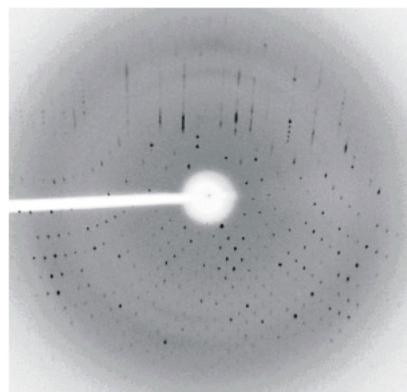
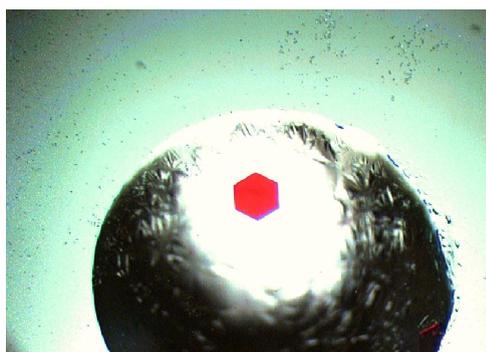
AMD was also shown to bind tightly and specifically to single-stranded DNA (Jovin & Wadkins, 1991, Jovin et al., 1996). The primary requirement is for guanine residues without need of a base-paired cytosine of an ordered structure. The fluorescent derivative of AMD, 7-aminoactinomycin D (7-AAMD), has the biological activity and dsDNA-recognition properties of AMD. Quantitative measurements of the changes in absorption and fluorescence of 7-AAMD upon interaction with a series of non self-complementary oligonucleotides were performed in the group of Prof. Jovin. Among the DNAs studied, the highest affinity was for the tetranucleotide sequence TAGT, which increases with the length of the oligomer. The drug to DNA ratio is 1:1, implying a single-stranded species.

The interactions of AMD with ssDNA may prove as important in understanding its biological mechanism as those involving dsDNA. It could be speculated that one mode of interaction of AMD in the inhibition of RNA polymerase involves binding in the domain in which strand separation occurs. Furthermore, AMD has been reported to be a potential inhibitor of human immunodeficiency virus type 1 (HIV-1) replication due to its ability to inhibit reverse transcription (Levin et al., 1998). In this case AMD inhibits the strand transfer step, in which the newly synthesized cDNA or (-)ssDNA is transferred from the donor RNA onto the acceptor RNA. Possibly, AMD acts through direct interaction with the (-)ssDNA, thereby blocking the basepairing capacity of this molecule.

In collaboration with the group of Prof. Jovin the crystallization of a series of single-stranded oligomers with 7-aminoactinomycin D was undertaken. The chosen oligomers contained the central sequence 5'-TAGT-3' with a varying number of flanking thymines. Previous tries in our department to crystallize a 7-AAMD-TAGT complex failed and the longer sequence 5'-TTAGTT-3' was chosen for the next attempt. The elucidation of its structure in complex with the drug will be discussed in this thesis.

3.2 FORMER WORK

In previous studies (E. Alexopoulos, Diplomarbeit 2001) crystals of the 7-AAMD-TTAGTT complex were obtained. The crystals grew at 20° C in 2-3 days from solutions containing 7-AAMD, TTAGTT, Gadolinium nitrate, ammonium sulfate, Na/K tartrate and a citrate buffer in a hanging drop setup. They showed a hexagonal symmetry (Fig. 3.3a).



a)

b)

Fig. 3.3 a) Native hexagonal crystals of the complex 7-AAMD-TTAGTT, b) Diffraction pattern of the native crystals.

The crystals could only be frozen with perfluorinated oil as a cryo protectant. Their diffraction pattern (Fig. 3.3b) at 100 K showed streaks in one direction, so no low temperature data could be collected. Nevertheless, a hexagonal cell could be determined ($a = b = 71.475 \text{ \AA}$ and $c = 109.317 \text{ \AA}$). Variation of the concentration or consistency of the salts or the buffers, or change of the temperature or seeding could not increase the quality of the crystals.

Thus, room temperature data was collected from a crystal mounted in a capillary on a SMART 6000 with Osmic mirrors and a Bruker CCD detector. The resolution was up to 2.6 \AA and the overall intensity rather low ($I/\sigma(I) = 4.85$). Investigation of the data with XPREP revealed a hexagonal symmetry with $P3_1/P3_2$ or $P3_121/P3_221$ as possible space groups. No anomalous signal of the gadolinium present in the crystallization liquor could be detected.

In order to incorporate heavy atoms in the crystal a new DNA oligomer was considered, in which the thymine at position 5 was exchanged by a bromouracil (5'-TTAG[Br⁵U]T-3').

3.3 CRYSTALLIZATION

The DNA oligomer samples were obtained from BIOTEZ. Stock solutions were prepared by dissolving them in water without further purification. 7-Aminoactinomycin D was purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany) and dissolved in water as a stock solution. The concentrations of both DNA and 7-AAMD were determined via UV absorption measurements.

Two different crystal forms of the 7-AAMD complex with the brominated sequence TTAG[Br⁵U]T were obtained, a hexagonal and an orthorhombic one. Although the hexagonal crystals were obtained first, their structure was determined only after solving the orthorhombic form. Therefore, in this chapter (3.3 Crystallization), both forms are described in chronological order. In the following chapters (3.4 and 3.5) the orthorhombic form is described first and then the hexagonal one.

3.3.1 THE HEXAGONAL CRYSTALS

Several trials of growing crystals of the brominated sequence TTAG[Br⁵U]T complexed with 7-AAMD using commercially available crystallization kits such as Hampton or JMB failed. The only successful method was cross seeding with crystals of the non-brominated complex. The initial crystals were obtained via the hanging drop method at 20 °C from solutions containing 2.3 mM 7-AAMD, 2.3 mM DNA (TTAGTT), 2.3 mM Gadolinium nitrate, 2.7 M ammonium sulfate, 0.05 M Na/K tartrate and 0.1 M Na citrate buffer (pH 5.6). Seeds of these crystals were transferred via cat whiskers to drops, which were equilibrated at 20 °C, containing 2.5 mM 7-AAMD, 2.5 mM brominated DNA (TTAG[Br⁵U]T), 2.5 mM Gadolinium nitrate, 2.7 M ammonium sulfate, 0.05 M Na/K tartrate and 0.1 M Na citrate buffer (pH 5.6). Within a period of 2-3 days hexagonal crystals appeared, which grew further to dimensions up to 0.3 mm³.

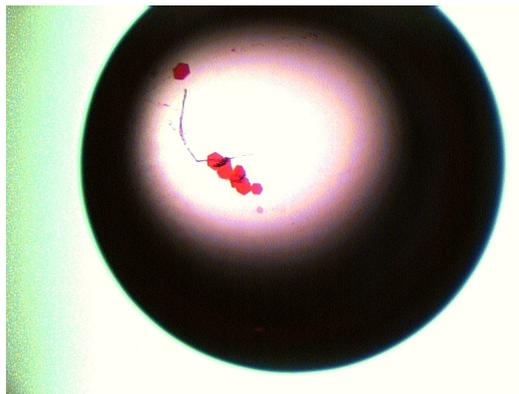


Fig. 3.4 Hexagonal crystals of the 7-AAMD-TTAG[Br⁵U]T complex obtained with cross seeding.

These crystals could be frozen without breaking by using perfluorinated oil as a cryo protectant. Indexing of the diffraction pattern was not conclusive, as two possible cells could be determined, depending on the reflections with $l = 2n + 1$, which are absent in the case of lower symmetry $P3_1/P3_2$. The first cell, which corresponds to the higher symmetry $P3_121/P3_221$, was similar to the one of the initial crystals:

$$a = b = 71.475 \text{ \AA} \text{ and } c = 109.317 \text{ \AA}$$

In the second case the c -axis was halved

$$a = b = 71.475 \text{ \AA} \text{ and } c = 54.94 \text{ \AA}$$

indicating the lower symmetry space groups $P3_1/P3_2$.

3.3.2 THE ORTHORHOMBIC CRYSTALS

Orthorhombic crystals were also obtained using the cross seeding method. The initial seeds were grown via the hanging drop method at 20 °C from solutions containing 2.3 mM 7-AAMD, 2.3 mM DNA (TTAGTT), 2.7 M ammonium sulfate, 0.05 M Na/K tartrate and 0.1 M Na citrate buffer (pH 5.6). They were transferred into drops of the same conditions but with a brominated DNA (TTAG[Br⁵U]T). The crystals obtained were long flat plates showing an orthorhombic habit.

The reason for the difference in geometry may lie in the fact that the initial crystals were about a year old, thus it is possible that orthorhombic seeds grew in the period of one year on the already existing hexagonal crystals.

Most of these crystals were two-dimensional and could not be measured. After several trials one of them showed a reasonable size in all three dimensions and could be frozen with perfluorinated oil as a cryoprotectant.



Fig. 3.5 Orthorhombic crystals of the 7-AAMD-TTAG[Br⁵U]T complex obtained with cross seeding.

3.4 THE ORTHORHOMBIC FORM

3.4.1 DATA COLLECTION

Table 3.2: Data collection statistics		
	Orthorhombic form	
wavelength source	1.5418 Å rotating anode	0.8110 Å X11
date (m.y)	05.02	06.02
crystallographic data		
<i>a</i> (Å)	51.556	39.453
<i>b</i> (Å)	70.885	51.586
<i>c</i> (Å)	39.661	70.975
space group	<i>C</i> 222 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
resolution (Å)	2.4	2.0
no. of unique reflections	6220	10296
$\langle I/\sigma(I) \rangle$	11.57 (3.04)	11.87 (6.33)
R_{merge}	0.045 (0.169)	0.032 (0.090)
completeness (%)	99.4 (99.1)	99.6 (98.8)
redundancy	1.80 (1.83)	1.83 (1.86)

Table 3.2 Data collection statistics for the orthorhombic form. Values in brackets refer to the outer 0.1 Å wide resolution shell.

Two low temperature datasets were collected. The first measurement was performed in-house on a Bruker SMART 6000 CCD detector to 2.4 Å resolution at a wavelength of 1.5418 Å (copper radiation) and a temperature of 100K. Data was collected in thin slices using an oscillation angle of 0.2° with a crystal-to-detector distance of 550 mm. A total of 6220 unique reflections were measured from a crystal of the size 0.2 × 0.2 × 0.1 mm³ mounted in a loop. Data integration and scaling was performed using the programs PROTEUM and SADABS (Bruker AXS, 2000). The cell determined was consistent with the orthorhombic symmetry having the following dimensions:

$$a = 51.556 \text{ \AA}, b = 70.886 \text{ \AA} \text{ and } c = 39.661 \text{ \AA}$$

XPREP was used to determine a C-centered lattice with the space group $C222_1$ due to the systematic absences. The anomalous signal detected is shown in Fig. 3.6. In general, an anomalous signal to noise ratio is significant starting from a value of 1.3, whereas a ratio of 1.0 is random. In this case, it is difficult to set a cutoff resolution for estimation of the F_A -values, as the anomalous signal to noise ratio is varying from 1.24 to 1.48 in the range 3.0-4.4 Å.

Anomalous signal/noise ratios (1.0 is random). The first line is based on input sigmas, the second on variances of F+ and F- (if not already averaged):

Inf - 8.0 - 6.0 - 5.0 - 4.6 - 4.4 - 4.2 - 4.0 - 3.8 - 3.6 - 3.4 - 3.2 - 3.0 A

1.78 1.74 1.69 1.54 1.31 1.37 1.48 1.25 1.24 1.33 1.30 1.45

50.8 Neighbors used on average for F+/F- local scaling
Rint(anom) = 0.0553 before and 0.0548 after local scaling

Fig. 3.6 Anomalous signal estimated by XPREP for the native data set.

A further data set of the same crystal was collected at EMBL, c/o DESY, Hamburg, on the X11 synchrotron beamline with an X-ray wavelength of 0.8110 Å using a MAR CCD detector. The data set was collected at 100 K with an oscillation angle of 0.5°. The resolution of the data (2.0 Å) was higher than for the in-house measurement, but the intensity was roughly the same. However, the cell determination with DENZO yielded a primitive setting with the cell axes:

$$a = 39.453 \text{ \AA}, b = 51.586 \text{ \AA} \text{ and } c = 70.975 \text{ \AA}$$

After integration and scaling of the data with DENZO and SCALEPACK XPREP suggested the primitive space group $P2_12_12_1$. The difference between this and the formerly found $C222_1$ are the reflections obeying:

$$h + k \neq 2n$$

In the case of $C222_1$ these reflections are systematically absent, whereas for $P2_12_12_1$ they are existent, but weaker, as they possess an $I/\sigma(I)$ of 8.8, while the rest of the reflections show an $I/\sigma(I)$ of 11.7.

The anomalous signal for the synchrotron data was slightly stronger than for the in-house data, nevertheless in general showing a similar behaviour to the in house data in the region of 2.4-3.2 Å (Fig. 3.7).

```
Anomalous signal/noise ratios (1.0 is random). The first line is based on
input sigmas, the second on variances of F+ and F- (if not already averaged):

Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.2 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.0 A

  2.01  1.53  1.57  1.37  1.35  1.25  1.37  1.51  1.56  1.43  1.16  1.11

109.7 Neighbors used on average for F+/F- local scaling
Rint(anom) = 0.0613 before and 0.0618 after local scaling
```

Fig. 3.7 Anomalous signal estimated by XPREP for the synchrotron data.

3.4.2 STRUCTURE SOLUTION

The two data sets were collected from the same crystal but at different wavelengths. As a first attempt the weak anomalous signal of the synchrotron data set was used to search for the bromine substructure in both space groups $P2_12_12_1$ and $C222_1$.

In the case of the primitive cell a SHELXD run of 1000 tries and with a minimum E value of 1.5 was started, searching for 8 bromine atoms at a high resolution cutoff of 3 Å. The correlation coefficients obtained for most of the trials were quite high (highest CC All/Weak: 45.62/31.32) suggesting that the structure was solved. A detailed examination showed that the solutions always included eight bromine sites, which correspond to eight strands per asymmetric unit (Fig.3.8).

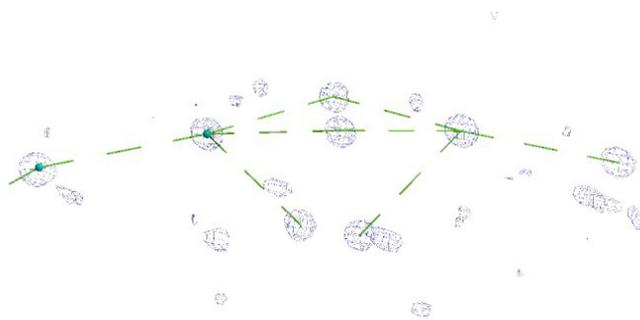


Fig. 3.8 One of the solutions containing 8 bromine sites found with SHELXD in space group $P2_12_12_1$.

The eight bromine atoms in all solutions showed two-fold non-crystallographic symmetry. Nevertheless, the attempts to superimpose the atoms of similar solutions failed, as the equivalent bromine positions differed up to 2 Å, making it impossible to choose the right solution.

As an alternative a *C*-centering was imposed for the synchrotron data. In a run of 1000 tries and with a minimum *E*-value of 1.3 SHELXD found four heavy atom sites, three of which were strongly occupied and one rather weak (Fig. 3.9) with high correlation coefficient values (CC All/Weak: 51.73/31.83).

BR01	1	0.344170	0.386147	0.210167	1.0000	0.2
BR02	1	0.376991	0.441841	-0.071189	0.8737	0.2
BR03	1	0.169273	0.511459	-0.066604	0.8134	0.2
BR04	1	0.215591	0.263664	-0.021148	0.5304	0.2

Fig. 3.9 The bromine positions found with SHELXD for the synchrotron data, assuming space group $C222_1$.

With these positions initial phases were calculated and afterwards used in DM (Cowtan, 1994) for solvent flattening and histogram matching. The whole resolution range and a solvent content of 60% were used. The resulting map was of a poor quality, but base stacking was visible and some peaks could be recognized as possible phosphates (Fig.3.10).

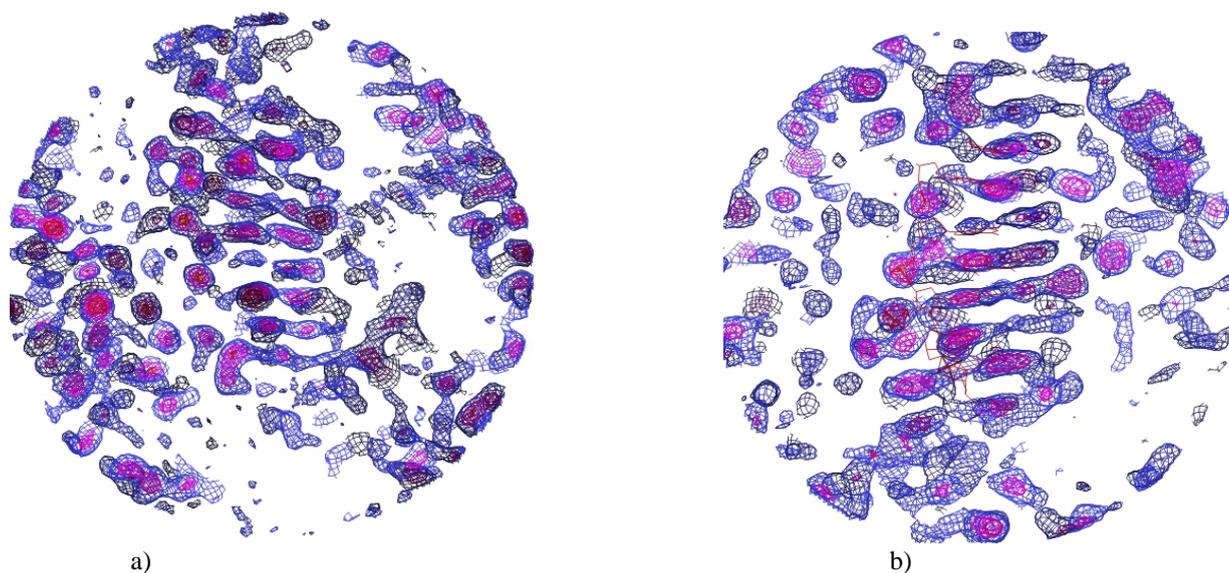


Fig. 3.10 a) First map obtained with DM, in which base stacking can be observed as well as red peaks, possibly representing phosphates b) First map obtained with DM with molecular modeling model fitted in density.

In order to obtain a better map a pseudo MAD experiment with both the in house and the synchrotron data was carried out. Although the two data sets were collected at wavelengths

not intended for MAD, they show a significant difference in the anomalous signal that could be used to calculate F_A values (Fig. 3.11).

Anomalous correlation coefficients (%) against previous datasets					
Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.2 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.0 Å					
81.9 80.1 78.4 68.2 55.9 45.6 47.9 45.4 29.7 14.2 20.3 9.3					
Set	f'	Rf'	f''	Rf''	after MAD fit
1	-1.500	0.0000	3.200	0.0262	synchrotron.sca
2	-0.700	0.0000	1.000	0.0837	inhouse.hkl

Fig. 3.11 Anomalous correlation coefficients, dispersive and anomalous scattering contributions f' , f'' of the two data sets collected over the resolution range (output of XPREP).

After local scaling of the merged in house and the merged synchrotron data a R_{merg} of 0.24 resulted. A R_{merg} value in the region of 20-25% indicates a good similarity between the data sets, but is nevertheless high enough to suggest that some change took place between the two measurements, maybe because of crystal damage during the measurements or the storing. The F_A values were calculated with the f' and f'' values from the Sasaki table (Sasaki, 1989) for the corresponding wavelengths. The existing bromine positions (Fig. 3.9) were used together with these data sets as an input for density modification in SHELXE. 20 cycles of density modification with 45% solvent were performed with SHELXE for both enantiomorphs. The mean estimated map CC was calculated for both cases and was slightly higher for the original hand indicating it as the correct structure (Fig. 3.12).

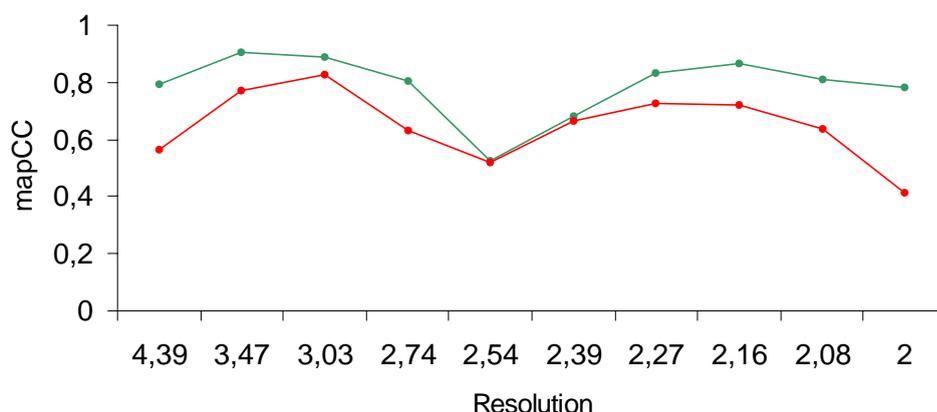


Fig. 3.12 Mean estimated mapCC from SHELXE after 20 runs of density modification against resolution (Å) for original (green) and inverted hand (red).

For this structure the *contrast* and the *connectivity* showed only slight differences for the two hands, being for the original somewhat higher (*contrast* = 0.390, *connectivity* = 0.867) than for the inverted (*contrast* = 0.335, *connectivity* = 0.846). The electron density map obtained was further modified with DM, with a solvent content of 30% and the whole resolution range. In this map one 2:1 DNA:7AAMD complex could be traced in the asymmetric unit by hand.

3.4.3 STRUCTURE REFINEMENT

The model was subjected to least-squares refinement against F^2 of the synchrotron data set with SHELXL97 (Sheldrick & Schneider, 1997) using a conjugate gradient algorithm. A set of reflections representing 5% of the overall reflections was chosen in thin shells to define an R_{free} set for cross validation (Brünger, 1992). The refinement stages are summarized in Table 3.3.

Action (No. of steps)	Atoms of complex	Water	Parameters	Restraints	R1 [$I > 2\sigma(I)$ / all]	Rfree [$I > 2\sigma(I)$ / all]
model building (10)	587	-	2351	2537	0.2871/0.2940	0.3179/0.3238
water addition (1)	587	21	2435	2558	0.2700/0.2764	0.3147/0.3232
anisotropic scaling (1)	587	21	2447	2592	0.2557/0.2615	0.3089/0.3152
bromine occupancy (1)	587	21	2448	2584	0.2512/0.2567	0.2952/0.3008
model (3)	572	15	2348	2491	0.2362/0.2422	0.2787/0.2856
hydrogen addition (2)	894	15	2400	2516	0.2357/0.2415	0.2887/0.2943

Table 3.3 Refinement progress in the orthorhombic case.

In the first 10 steps the complex was built with XFIT (McRee, 1999) by tracing the map by hand and using parts of the molecular dynamics model as templates. In order to stabilize the refinement a stepwise increase of the resolution range was applied. The high-resolution limit was set to 3 Å, so that still enough data were available relative to the number of parameters. The geometrical 1,3- and 1,2-restraints of Parkinson et al. for DNA were used. In the case of the bromouracil moiety they were generated with SHELXPRO using the structure deposited in the CSD. The geometrical 1,3- and 1,2-restraints for the 7-AAMD molecule were generated with SHELXPRO using the atomic coordinates of the high-resolution crystal structure of Actinomycin D (1A7Y.pdb). Furthermore, planarity and chiral volume restraints

were defined. The diffuse solvent region was accounted for by a two parameter model after Babinet's principle (Moews & Kretsinger, 1975). An initial R_{work} of 0.29 ($R_{\text{free}} = 0.32$) resulted. As a next step anti-bumping restraints were used and water molecules were added manually in the map, selecting among the highest electron-density peaks those making reasonable contacts. The approximately spherical shape of the difference electron density ($F_o - F_c$) corresponding to water molecules was confirmed with the σ -weighted map. The model was completed to a R_{work} of 0.27 and $R_{\text{free}} = 0.31$.

Taking into account the long needle-like shape of the crystals an anisotropic scaling correction was considered. In this procedure twelve anisotropic scaling parameters are refined for the otherwise isotropic structure (Usón et al., 1999). This resulted in R-factors of $R_{\text{work}} = 0.26$ and $R_{\text{free}} = 0.31$.

During the refinement it was noticed that the B -values of the bromine atoms were higher compared to the other atoms of the corresponding base. The refinement of their occupancy taking into account the absorption at the current wavelength ($f' = -1.5 \text{ e}$, $f'' = 3.2 \text{ e}$) resulted in a value of 66.45%. This x-ray-induced debromination is a consequence of bromine cleavage resulting from direct photodissociation and/or the action of free electrons (Enifar et al., 2002).

The inclusion of H atoms at geometrically calculated positions and their refinement using a riding model did not lead to any improvement, but increased the R_{free} -values and was therefore rejected.

The final refinement statistics are shown in Table 3.4.

Unit cell parameters (Å)	$a = 51.586 (0.052)$ $b = 70.975 (0.071)$ $c = 39.453 (0.039)$
Space group	$C222_1$
No. of unique reflections	10296
Resolution range	20-2.0
$R_{\text{work}} (I > 4\sigma)$	0.2362
$R_{\text{free}} (I > 4\sigma)$	0.2787
No. of water molecules	13
Data/restraints/parameter	4870/2493/2356
R.m.s. deviations from idealized geometry	
Bond lengths (Å)	0.006
1,3-distances (Å)	0.019
Zero chiral volumes (Å ³)	0.000
Non zero chiral volumes (Å ³)	0.004
Distances from restraint planes (Å)	0.205
Mean B factor (Å²)	
DNA atoms	21.9
7-AAMD atoms	15.2

Table 3.4 Final refinement statistics of the orthorhombic structure

3.5 THE HEXAGONAL FORM

3.5.1 DATA COLLECTION

From a single crystal of the hexagonal form with the dimensions $0.3 \times 0.3 \times 0.01 \text{ mm}^3$ five data sets at different wavelengths were collected at low temperature. Initially the big cell

$$a = b = 71.475 \text{ Å and } c = 109.317 \text{ Å}$$

was chosen for the integration and scaling. In all cases an acceptable space group could not be determined by XPREP as the $\langle |E^2 - 1| \rangle$ statistics suggested a centrosymmetric structure (standard values: 0.74 for non-centrosymmetric and 0.97 for centrosymmetric). This problem is rather characteristic for DNA structures. As they show a strong stacking of their bases, there is no statistical distribution of the atoms, so the $\langle |E^2 - 1| \rangle$ statistics fail. The space group chosen was the one found in the native crystals $P3_121/P3_221$.

Table 3.5: Data collection statistics					
	Hexagonal form				
Wavelength (Å) Source	1.5418 Å in house	0.8463 Å BW7B	0.9196 Å peak BW7A	0.9204 Å inflection BW7A	0.9050 Å high-remote BW7A
Date	11.01	12.01	07.02	07.02	07.02
Cell dimensions					
a (Å)	71.230	71.475	71.216	71.168	71.222
b (Å)	71.230	71.475	71.216	71.168	71.222
c (Å)	108.490	109.317	108.408	108.347	108.400
Space group	$P3_121/P3_221$	$P3_121/P3_221$	$P3_121/P3_221$	$P3_121/P3_221$	$P3_121/P3_221$
Resolution (Å)	2.5	2.5	2.4	2.4	2.5
Unique reflections	11394	11493	12988	12964	11527
$\langle I/\sigma(I) \rangle$	5.31 (7.5)	7.98 (8.2)	6.03 (4.4)	6.53 (3.5)	6.57 (5.3)
R_{merge}	0.113 (0.519)	0.072 (0.284)	0.088 (0.728)	0.075 (0.658)	0.080 (0.480)
completeness (%)	99.3 (99.8)	99.0 (92.7)	98.9 (92.1)	99.1 (95.2)	99.0 (93.3)
redundancy (outer shell)	17.0 (16.4)	10.9 (9.8)	30.6 (3.1)	30.2 (2.9)	30.9 (2.8)

Table 3.5 Data collection statistics for the hexagonal form. Values in brackets refer to the outer 0.1 Å resolution shell.

The first measurement was performed in house on a mar345 image plate detector with Bruker rotating anode and Osmic mirror system to 2.5 Å resolution at a wavelength of 1.5418 Å (copper radiation) and a temperature of 100 K. A thick slicing mode was used with an oscillation angle of 0.5°. The cell determination with DENZO resulted in a hexagonal geometry with Laue group -3. The data was processed with DENZO and SCALEPACK leading to a total of 11394 unique reflections with an $\langle I/\sigma(I) \rangle$ of 5.31 and a mosaicity of 0.8°. This data set was intended as a first in-house test of the crystal and has the lowest intensity and redundancy resulting in a higher R_{merge} than the synchrotron data.

The crystal was then frozen and data were collected at EMBL, c/o DESY, Hamburg, on the BW7B synchrotron beamline with an X-ray wavelength of 0.8463 Å using a mar345 image plate detector. Using an oscillation angle of 0.5°, within a range of 49° a data set to 2.5 Å resolution was collected. Data processing was performed with DENZO and SCALEPACK resulting in 11493 unique reflections with an $\langle I/\sigma(I) \rangle$ of 7.98. The completeness was 99% and the R_{merge} 0.072. These data are stronger than the in-house data owing to the stronger X-ray source.

After the solution of the orthorhombic structure, a MAD data set of the hexagonal crystal was collected at EMBL, c/o DESY, Hamburg, on the BW7A synchrotron beamline and using a mar CCD unsupported-m165 detector. A fluorescence spectrum was first acquired in order to accurately locate the bromine *K* edge (Fig. 3.13).

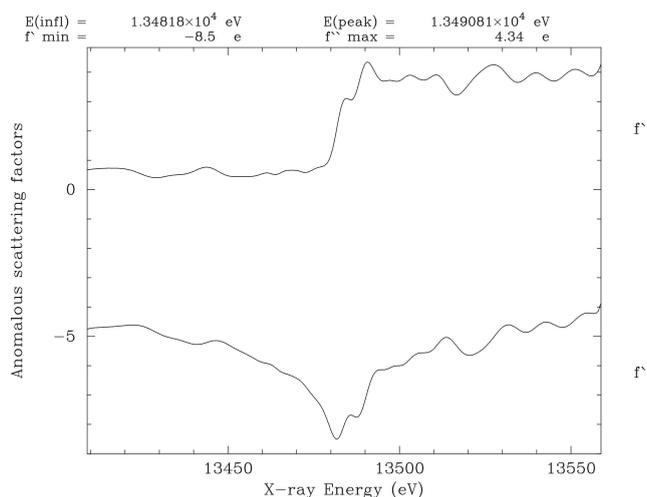


Fig. 3.13 X-ray fluorescence spectrum for the complex 7-AAMD-TTAG[Br⁵U]T obtained with the program CHOOCH (Evans & Pettifor, 2001). The spectrum starts with a value of about 0.6 e for the f'' and -5 e for f' . The f' curve decreases continuously from 13450 eV on and reaches a minimum of -8.5 e at 13482 eV corresponding to 0.9204 Å (inflection point). The increase of the f'' curve starts at 13475 eV until the absorption edge of 4.34 e of the bromine occurs at 13491 eV corresponding to 0.919 Å (peak point). Just before the actual peak is reached a light shoulder is observed at about 13482 eV.

Three data sets were measured at the wavelengths 0.9196 Å (peak), 0.9204 Å (inflection) and 0.9050 Å (high-energy remote). The data was collected at 100 K and a thick slicing mode of 0.5° was used. Integration and scaling were performed with DENZO and SCALEPACK. The completeness and the redundancy were acceptable for all three measurements, but the intensity was rather low (Table 3.5).

3.5.2 STRUCTURE SOLUTION

Due to the change of symmetry, it was not proven that the structure of the orthorhombic crystals was the same as in the native hexagonal crystal. As a proof the hexagonal brominated structure had to be elucidated as well. This was done using two different methods. The solution of the orthorhombic form suggested the use of MR techniques with this structure as a searching model. On the other hand, the MAD data set collected from the existing hexagonal crystal could also be used to solve the second form.

3.5.2.1 MOLECULAR REPLACEMENT

The previously solved and refined structure of the 7-AAMD complex affords a suitable model for molecular replacement at low resolution. The data used was at that point only the available synchrotron data measured on BW7B.

Table 3.6: Molecular replacement trials						
Input						
Cell axis γ (Å)	108.1	108.1	54.94	54.94	54.94	108.1
Space group	$P3_121$	$P3_221$	$P3_121$	$P3_221$	$P3_1$	$P3_221$
Program used	COMO	COMO	COMO	COMO	COMO	COMO
Resolution (Å)	10-3.5	10-3.5	10-2.5	10-2.5	10-2.5	10-2.5
No. copies searched for	4	4	2	2	4	dimer
Model used	$C222_1$ sol.	$C222_1$ sol.	$C222_1$ sol.	$C222_1$ sol.	$C222_1$ sol.	$P3_1$ dimer
Results						
No. copies found	2	2	1	1	4	2 dimers
CC-values	mol1: 0.496 mol2: 0.515	mol1: 0.254 mol2: 0.441	mol1: 0.564	mol1: 0.524	mol1: 0.531 mol2: 0.572 mol3: 0.601	mol1: 0.596 mol2: 0.650
R-values	mol1: 0.549 mol2: 0.542	mol1: 0.504 mol2: 0.590	mol1: 0.511	mol1: 0.536	mol1: 0.507 mol2: 0.509 mol3: 0.490	mol1: 0.505 mol2: 0.485

Table 3.6 Molecular replacement trials in the hexagonal case.

In the first attempts molecular replacement was tried in the space groups $P3_121$ and $P3_221$ with the big cell ($a = b = 71.475$ Å and $c = 109.317$ Å). The limited six-dimensional search for four copies around the rotation function maxima in a resolution range of 15-3.5 Å with the program COMO gave no satisfactory results obtaining only two molecules in the asymmetric unit in both cases. Further copies were rejected due to sterical hindrance in the packing. Neither MR with the smaller cell ($a = b = 71.475$ Å and $c = 54.94$ Å) was successful.

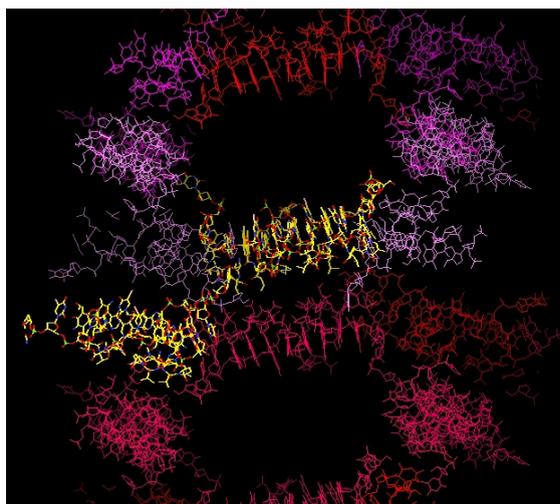


Fig. 3.14 Packing plot of the COMO solution in space group $P3_221$.

A detailed examination of the packing plot in the big cell for space group $P3_221$ showed that no steric clashes are observed between the symmetry equivalent molecules, but that there are empty spaces in the crystal packing, too big to be solvent channels. The keys are the two-

fold axes running along a positioned in the middle of these gaps making it impossible for non-symmetrical molecules to occupy the void space. Apparently the problem does not lie in the rotation function that seems to lead to a correct orientation of the fragment, but in the translation search.

In order to override this symmetry problem molecular replacement was tried in a lower symmetry space group, which lacked the two-fold axis. The maximal non-isomorphic subgroup for $P3_221$ is $P3_2$. By taking the halving of the c -axis into account $P3_2$ was transformed to its maximal isomorphic subgroup of lowest index $P3_1$. The total cell transformation was performed with the matrix

$$\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.5 \end{matrix}$$

The search was performed with the program COMO in the space group $P3_1$ for a resolution range of 10-2.5 Å. The four copies found showed a reasonable packing (Fig. 3.15)

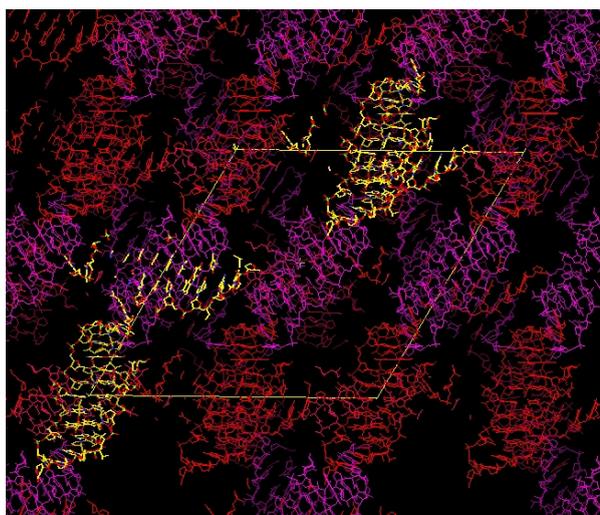


Fig. 3.15 Packing plot of the solution in $P3_1$ with COMO.

The values achieved for the correlation coefficient and the R-factor were reasonable (CC = 0.601, R = 0.490). The SHELXL refinement with this model using the same restraints as in the orthorhombic case resulted in unusually high R values ($R_{\text{work}} = 0.42$, $R_{\text{free}} = 0.54$), which in the case of an almost complete structure is often a symptom for twinning. The incorporation of the two-fold twin matrix

$$\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{array}$$

in the refinement decreased the R values to $R_{\text{work}} = 0.38$ and $R_{\text{free}} = 0.46$. The BASF value converged to 0.50180. This suggests that either this is a perfect twin with two equally occupied domains or the correct space group is the higher symmetric $P3_221$ and the two-fold axis is a real crystallographic axis.

A molecular replacement attempt with COMO using two molecules of the $P3_1$ model as a search fragment in the space group $P3_221$ was successful. Four copies were found with $CC = 0.650$ and $R = 0.485$ and a reasonable packing (Fig. 3.16). The refinement values in $P3_221$ were $R_{\text{work}} = 0.26$ and $R_{\text{free}} = 0.30$.

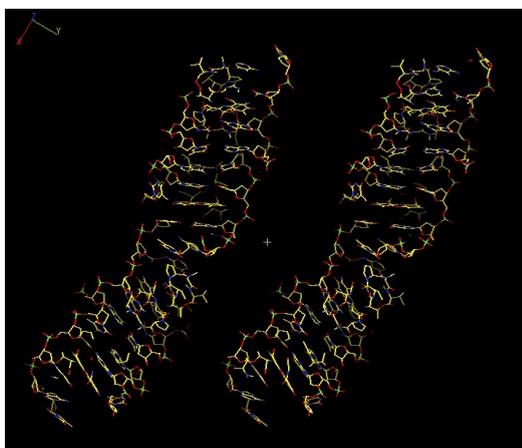


Fig. 3.16 The four copies found with COMO in space group $P3_221$.

3.5.2.2 MAD

The MAD experiment was performed with all four synchrotron data sets of the hexagonal crystal. The F_A estimation combining the four data sets was done with XPREP choosing the space group $P3_221$ in the big cell, as it hinted by the MR. The unmerged data sets were read in XPREP and scaled onto each other with the high-energy remote at 0.9050 Å as a reference. The inflection as well as the data set from BW7B had to be reoriented with the matrix

$$\begin{array}{ccc} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{array}$$

The R_{merg} values of the scaled data sets were low suggesting a good similarity (Table 3.7).

	0.8463 Å BW7B	0.9196 Å peak	0.9204 Å inflection
0.9050 Å remote	$R_{\text{merg}} = 0.0267$	$R_{\text{merg}} = 0.0363$	$R_{\text{merg}} = 0.0378$

Table 3.7 The R-values of the scaled data sets.

The f' and f'' values of the Sasaki table corresponding to the wavelengths were used as initial values and refined in XPREP (Table 3.8).

	initial values		refined values	
	f'	f''	F'	f''
0.8463 Å BW7B	-1.50	3.20	0.50	2.46
0.9050 Å remote	-3.33	3.70	-4.78	2.95
0.9204 Å inflection	-8.50	3.80	-7.75	4.50
0.9196 Å peak	-6.14	4.34	-5.54	4.55

Table 3.8 Initial and refined f' and f'' values used for the calculation of the F_A values.

After 113 tries the best SHELXD solution contained sixteen heavy atom sites (Fig. 3.17) and had a correlation coefficient of All/Weak 59.18/49.33.

BR01	1	0.586182	0.879295	0.026943	1.0000	0.2
BR02	1	0.953247	1.025162	-0.178471	0.9375	0.2
BR03	1	0.574982	0.704529	0.117321	0.8956	0.2
BR04	1	0.605103	1.114471	0.040149	0.8941	0.2
BR05	1	0.802155	0.931351	-0.138813	0.8899	0.2
BR06	1	0.703278	1.124741	-0.192341	0.8087	0.2
BR07	1	0.438744	0.966354	0.010296	0.8044	0.2
BR08	1	0.406456	0.518364	0.132602	0.7763	0.2
BR09	1	0.625877	1.099442	-0.025974	0.7590	0.2
BR10	1	0.429459	1.133156	0.051975	0.6765	0.2
BR11	1	0.640007	0.929665	-0.046864	0.6468	0.2
BR12	1	0.475021	0.937439	-0.004322	0.6443	0.2
BR13	1	0.974594	0.879044	-0.206814	0.6301	0.2
BR14	1	0.381447	1.079300	-0.017636	0.6204	0.2
BR15	1	0.778992	0.857742	-0.224980	0.6111	0.2
BR16	1	0.570320	0.531708	0.155964	0.5354	0.2

Fig. 3.17 Bromine sites found with SHELXD.

30 cycles of density modification assuming 30% solvent were performed with SHELXE for both enantiomorphs $P3_121$ and $P3_221$. The mean estimated map CC was calculated for both cases and was highest for $P3_221$ (Fig. 3.18) identifying it as the correct space group.

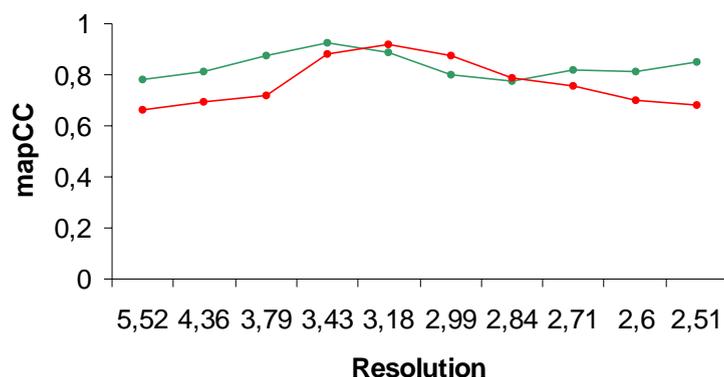


Fig. 3.18 Mean estimated map CC from SHELXE after 30 runs of density modification against resolution (Å) for $P3_121$ (red) and $P3_221$ (green).

Both the *contrast* and the *connectivity* were lower for $P3_121$ (*contrast* = 0.321, *connectivity* = 0.843) than for $P3_221$ (*contrast* = 0.382, *connectivity* = 0.885). Four copies of the orthorhombic structure could be fitted in the electron density map obtained (Fig. 3.19) and further bases could be traced by hand.

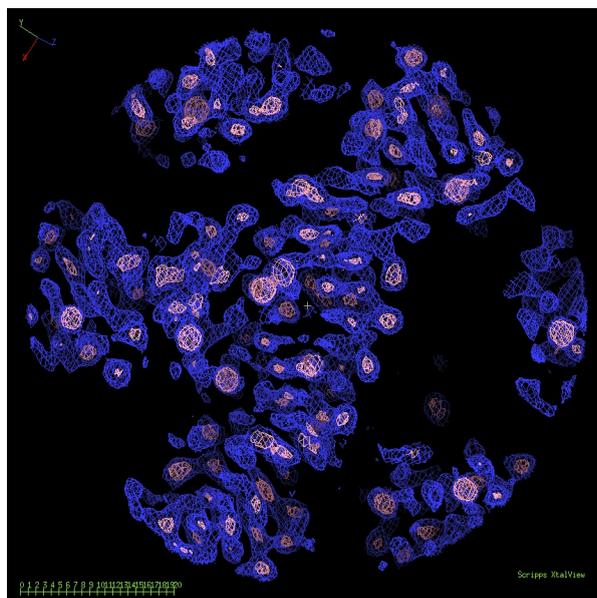


Fig. 3.19 Electron density map obtained with SHELXE from MAD data.

3.5.3 STRUCTURE REFINEMENT

The model was refined against F^2 with SHELXL97 using a conjugate gradient algorithm. The high-energy remote data at BW7A was used as the reference as they resulted the best R-values of all data sets. A set of reflections representing 5% of the overall reflections was chosen in thin shells to define an R_{free} set (Brünger, 1992) for cross validation. For DNA

structures it is more common to choose the free reflections randomly. Nevertheless, in this case the sixteen bromine atoms in the asymmetric unit suggest the existence of NCS, which justifies a thin shell method. The refinement stages are summarized in Table 3.9.

Action (No. of steps)	Atoms of complex	Water	Parameters	Restraints	R1 [$I > 2\sigma(I)$ / all]	Rfree [$I > 2\sigma(I)$ / all]
model building (3)	2240	-	8963	17273	0.2372/0.2491	0.2973/0.3084
HOPE, SIMU, BUMP (5)	2232	-	8943	17011	0.2669/0.2786	0.3058/0.3156
water addition (1)	2268	42	9112	17013	0.2534/0.2721	0.2916/0.3115
hydrogen addition (1)	3018	42	9108	17015	0.2533/0.2718	0.2944/0.3131
bromine occupancy (1)	2232	14	8996	17002	0.2576/0.2760	0.2962/0.3163

Table 3.9 Refinement progress in the hexagonal case.

In the first three steps the complex was built in XFIT by fitting four copies of the orthorhombic structure in the electron density. The same geometrical 1,3- and 1,2-restraints were used for the DNA and the 7-AAMD as in the orthorhombic case. Furthermore, planarity and chiral volume restraints were defined. The diffuse solvent region was accounted for by a two parameter model after Babinet's principle (Moews & Kretsinger, 1975). Local non-crystallographic symmetry restraints were applied to similar parts of the four complexes using the instruction NCSY. There resulted an initial R_{work} of 0.24 ($R_{\text{free}} = 0.30$). In the next steps anti-bumping and similarity restraints were used and an anisotropic scaling correction was applied. The increase of the R-values ($R_{\text{work}} = 0.27$, $R_{\text{free}} = 0.31$) in comparison to the model building steps can be justified by the decrease of the restraints.

The addition of water molecules in the map improved the R_{work} to 0.25 and the R_{free} to 0.30.

The refinement of H atoms in geometrically calculated positions by use of a riding model did not lead to any improvement and was abandoned.

The occupancy of the bromine atoms was refined to 53% by taking into account their absorption at the particular wavelength ($f' = -3.33$ e, $f'' = 3.70$ e). The refinement of the model against each of the collected data sets showed a successive crystal damage (Table 3.10).

Beamline	in-house mar	BW7B	BW7A		
Date (m.y)	11.01	12.01		7.02	
			peak	inflection	remote
Wavelength(Å)	1.54	0.8463	0.9196	0.9204	0.905
Sasaki f'/f''	-0.77/1.28	-1.62/3.27	-6.14/3.82	-7.06/3.82	-3.33/3.70
Bromine occupancy	0.83	0.69	0.59	0.50	0.53

Table 3.10 Occupancy decrease of bromine sites.

The final refinement statistics in the hexagonal case are shown in Table 3.11.

Unit cell parameters (Å)	$a = b = 71.487 (0.071)$ $c = 109.160 (0.109)$
Space group	$P3_221$
No. of unique reflections	11527
Resolution range	20-2.5
$R_{\text{factor}} (I > 4\sigma)$	0.2576
$R_{\text{free}} (I > 4\sigma)$	0.2962
No. of water molecules	14
Data/restraints/parameter	10926/17002/8996
R.m.s. deviations from idealized geometry	
Bond lengths (Å)	0.005
1,3-distances (Å)	0.015
Zero chiral volumes (Å ³)	0.000
Non zero chiral volumes (Å ³)	0.003
Distances from restraint planes (Å)	0.159
Mean B factor (Å²)	
DNA atoms	30.1
7-AAMD atoms	49.8

Table 3.11 Final refinement statistics for the hexagonal structure.

3.6 STRUCTURE OF THE 7-AAMD COMPLEX

3.6.1 OVERALL STRUCTURE

The structure described is based on the orthorhombic form. The hexagonal form has four copies of the complex ($2 \times 7\text{-AAMD}$, $4 \times \text{TTAG}[\text{Br}^5\text{U}]\text{T}$) in the asymmetric unit and shows slight differences, which will be discussed later.

The drug to strand stoichiometry is 2:4. The four conformationally different single strands pair partially to form two stacked antiparallel double strands. In each one of these double

strands an actinomycin molecule intercalates with the phenoxazine ring inserted between a G-G step (Fig. 3.20).

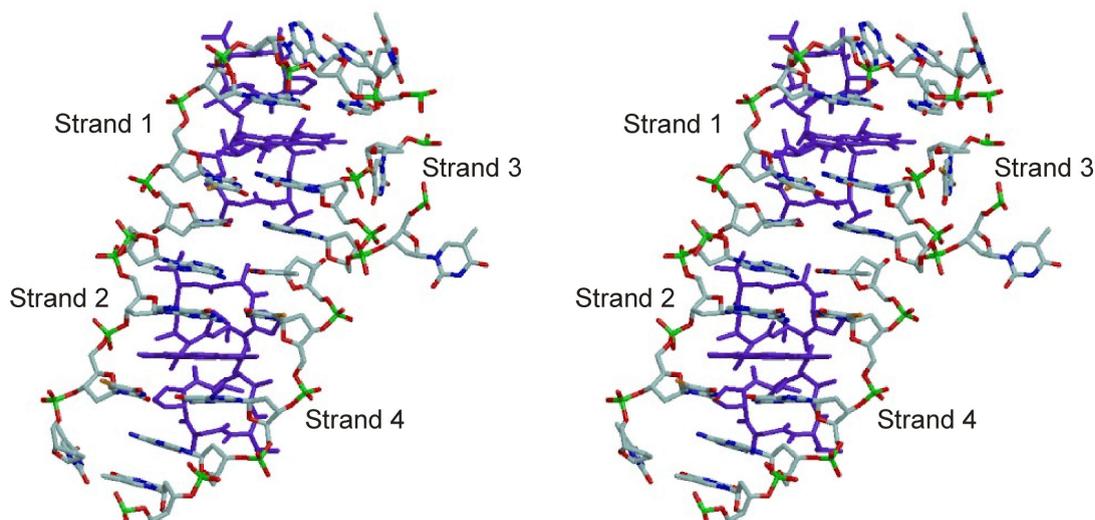


Fig. 3.20 Stereo view of the complex. The actinomycin molecules intercalating with the single strands are shown in lilac.

The first double strand, shown in Fig. 3.21, is formed by strands **1** and **3**. The single strands are connected with two stacked base pairs, T106-A303 and BrU105-G304.

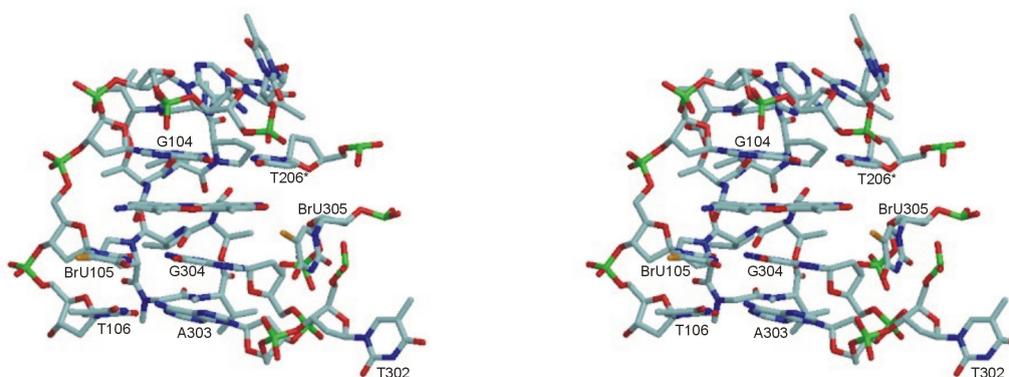


Fig. 3.21 Stereo view of the upper part of the complex consisting of strands **1** and **3**, intercalating with actinomycin **1**.

The latter shows an unusual GU wobble pairing in which the nitrogen N1 and the oxygen O6 of the guanine are hydrogen bonded to the oxygen O3 and the nitrogen N3 of the pyrimidine (Fig. 3.22 a). The stacking continues with the phenoxazine ring of the actinomycin and the guanine G104, which pairs with thymine T206* of a symmetry related molecule, again in a wobble mode (Fig. 3.22 b).

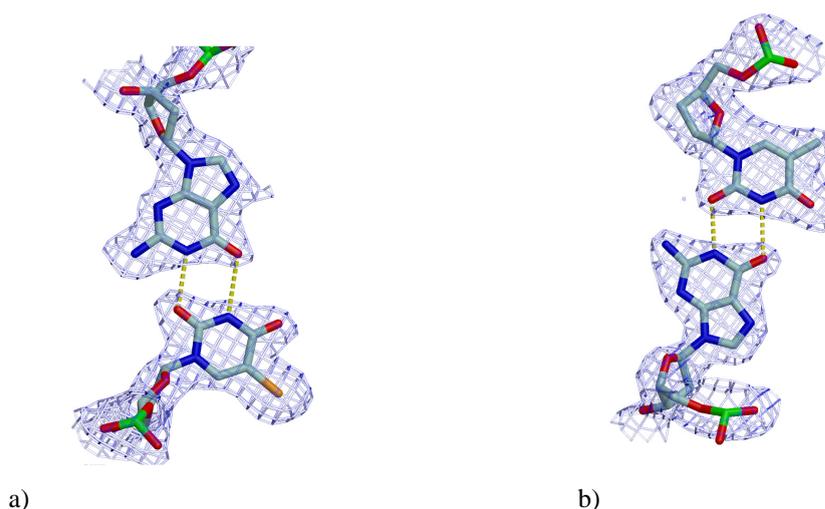


Fig. 3.22 a) The mismatched base pair G304-BrU105. Interaction takes place in a wobble mode with hydrogen bonding between the N1 and the O6 of the guanine with the O2 and the N3 of the bromouracil respectively, b) The mismatched base pair G104-T206*. The interaction is the same as in the G-BrU case.

In strand **1** the phosphate backbone performs a sharp turn after base G104 with a torsion angle $\delta_{G104} = 137.7^\circ$. The adenine A103 and the thymine T102 and T101 do not participate in the stacking. On the contrary, the angles of their planes to that of the guanine G104 are 65.0° , 27.8° and 60.7° respectively. In this way they can interact with symmetry related strands. Interestingly, the two thymine (T101 and T102) participate in T*A*T triplets, whereas the adenine (A303) forms an A-T pair. The remaining two bases of strand **3** are not involved in pairing. Bromouracil BrU305 is positioned perpendicular to the phenoxazone ring in a looped out fashion at an angle of 98.1° to the ring (Fig. 3.23). The bromine atom points to the methyl group C15 of the chromophore. A similar pattern of bases looping out has been described previously in double stranded DNA complexes with actinomycin (Chen *et al*, 2002).

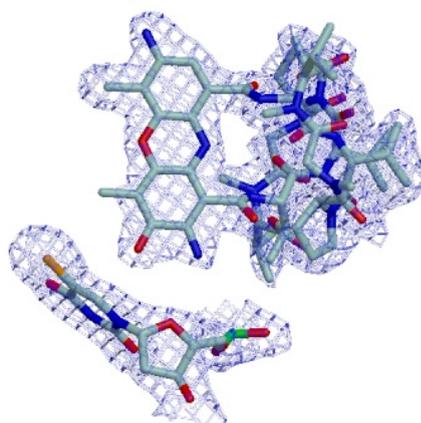


Fig. 3.23 π -interaction of BrU305 to the phenoxazone ring of actinomycin molecule **1**.

The thymine T302 interacts with its symmetry equivalent base T302* of a neighboring molecule (Fig. 3.24). The first and last thymine of **3** are missing in the experimental electron density map.

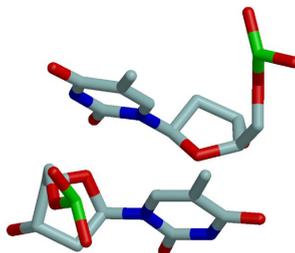


Fig. 3.24 Interaction of thymine 302 with its symmetry equivalent T302*.

The second part of the complex, consisting of strands **2** and **4**, is shown in Figure 3.25.

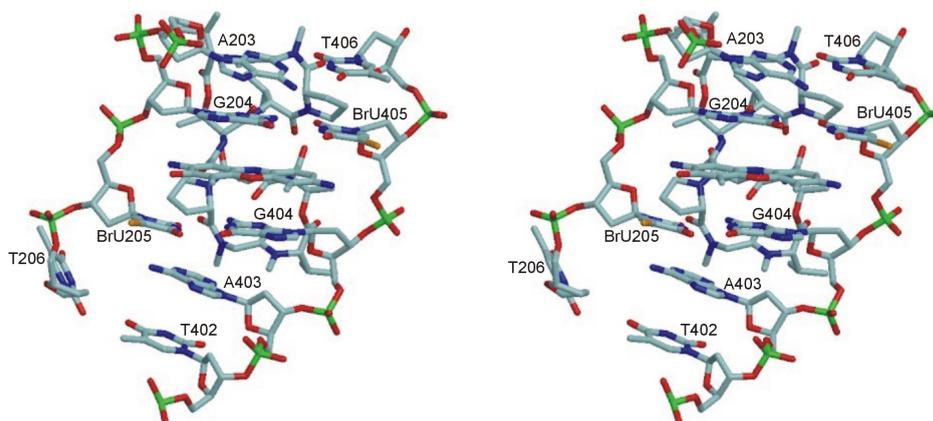


Fig. 3.25 Stereo view of the lower part of the complex, consisting of strands **2** and **4** interacting with actinomycin molecule **2**.

The strands interact via three hydrogen-bonded base pairs. A Watson-Crick A203-T406 pair stacks with the T106-A303 pair of the first part of the complex. Two G-BrU steps follow, sandwiching the chromophore group of the second actinomycin. The pairing is again of a GT wobble mode with both bromouracils participating. The remaining bases of strand **4** interact with symmetry related complexes. The adenine A403 participates in the T*A*T triplet mentioned above.

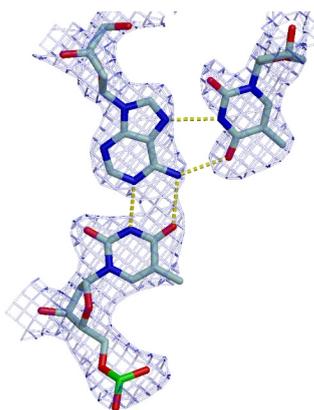


Fig. 3.26 The T*A*T triplet between the bases A403, T102* and T101*. The triplet involves Watson-Crick pairing between thymine 102* and the adenine (N3, O4 of T102* to N1, N6 of A403 respectively), and Hoogsten interactions between thymine 101* and the adenine (N3, O4 of T101* to N7, N6 of A403 respectively).

This T*A*T triplet (Fig. 3.26) consists of bases A403, T102* and T101* from three different strands with thymine T102* precedent from a symmetry related complex beneath, so long stacks of complexes are formed in the crystal. These stacks are cross linked mainly through the last base T101* of the T*A*T triplet, as well as the thymine-thymine stacking of T302. The thymine T402 pairs with A103* of the same stack. In strand **2** the thymine T206 is turned away to pair with the G104*, which again bridge over to a neighboring stack. Two bases of **2** and one base of **4** are missing in the experimental density. The missing bases in all four strands always flank thymines not involved in pairing. This could result in high flexibility and possible disorder, so that the map is difficult to trace.

3.6.2 CONFORMATION

All four strands are distorted in their structure (Fig. 3.27), the most distorted being strand **3**.

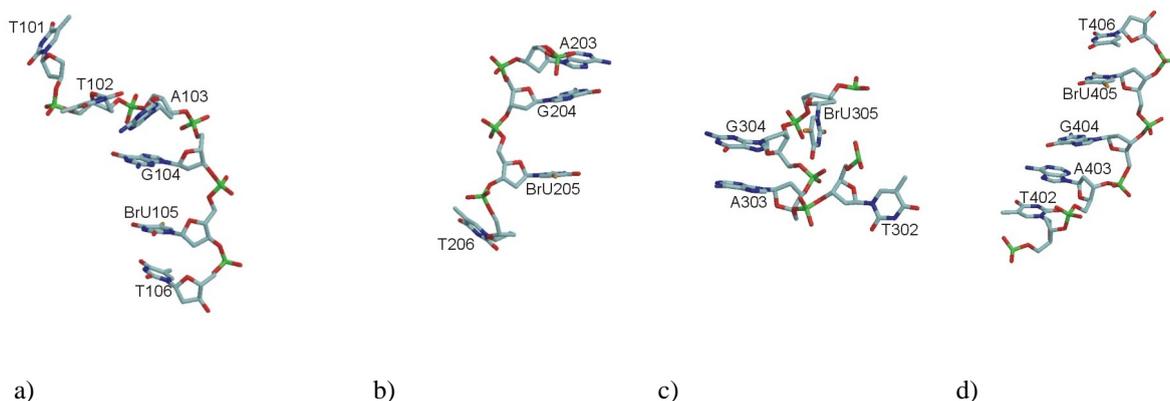


Fig. 3.27 Strands **1** (a), **2** (b), **3** (c) and **4** (d).

This flexibility is more characteristic for RNA than for DNA structures. In the dsDNA structure proposed by Watson and Crick the two strands formed a double helix in which the bases paired by complementary hydrogen bonding in the helix interior, the sugar-phosphate backbones extended along the outside and the strands ran in antiparallel directions. There proved to be several possible conformations of nucleic acid fibers (A-form, B-form, Z-form) depending on the environment. The structure of the double strands of the 7-AAMD complex does not correspond to any of these forms. The maximal number of bases of one single strand participating in pairing with just one other strand is three (strands **2** and **4**). The remaining bases either pair with symmetry equivalent bases corresponding to different strands or are so flexible that they are not visible in the density map. Therefore the structural parameters of rotation and translation along a regular double stranded helix cannot be used in this case. The TTAG[Br⁵U]T oligomers can only be described on their own by the seven torsion angles specifying the conformation of each nucleotide in a polynucleotide chain (Fig. 3.28).

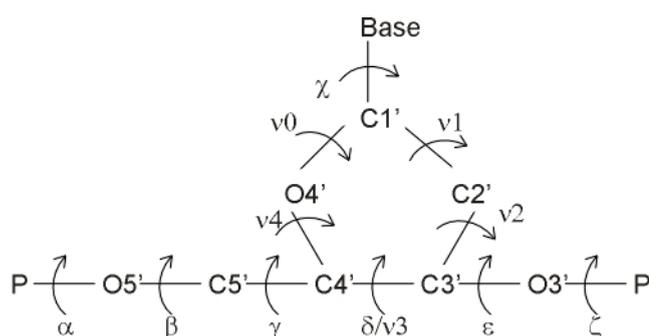


Fig. 3.28 The torsion angles specifying the conformation of a nucleotide. The six angles α , β , γ , δ , ϵ and ζ characterize the sugar-phosphate backbone, while the angle χ describes the orientation of the base relative to the sugar. The sugar conformation is defined by five torsion angles ν_0 , ν_1 , ν_2 , ν_3 and ν_4 .

These parameters show specific conformational preferences in nucleosides and nucleotides. The phosphodiester linkage rotation angles (α and ζ) favor the g^+ , g^+ (or g^- , g^-) conformation. Due to the minimization of the P-C4' repulsions the preferred value for the β angle is 180° . The tendency for the rotamers around C5'-C4' (γ) is $g^+ > t > g^-$, whereas for the ϵ torsion angle the g^+ conformation is the least favorite due to steric repulsions between P and C4' or C2' (tendency $t > g^-$). The torsion angle δ is directly related to the sugar conformation. It specifies the same torsion angle as ν_3 but has a different magnitude ($\delta = \nu_3 + 125^\circ$). The preferred values for the glycosidic angle χ are different for purines and pyrimidines. For purines, the syn ($\chi = 240^\circ$) and anti ($\chi = 60^\circ$) conformations are

comparable in energy, whereas for pyrimidines the anti conformation is energetically lower than the syn. Small differences can be observed between the A-, B- and Z-DNA forms.

As regards the ribose and deoxyribose there are five torsion angles defined for the five membered rings: ν_0 , ν_1 , ν_2 , ν_3 and ν_4 . Nevertheless, more common is the specification of the conformation by the atoms located out of the ring plane. For DNA and RNA it is C3'-endo and C2'-endo.

The values of the angles as well as the sugar conformations for the orthorhombic structure are shown in table A (Appendix). The distortion of the strands can be expressed by the deviation from the standard values. The conformations are strongly conserved for β , γ and ϵ , but vary for the bonds around the phosphorus. The bases of the strands are all anti with glycosidic angles χ varying from -71° to -167° . The sugar conformations are very different from the standard angles. For the bromouracils the preferred conformation is the unusual O4'-endo, C4'-exo. It also occurs for A103, T106 and T406. The remaining residues prefer the standard C2'-endo, C3'-exo.

3.6.3 INTERACTION WITH 7-AAMD

The interaction with the drug is mainly via stacking of the actinomycin phenoxazone ring with the two guanine residues of the ssDNAs. The guanine rings overlap dominantly on the phenoxazone whereas the thymine rings show very little overlap (Fig. 3.29).

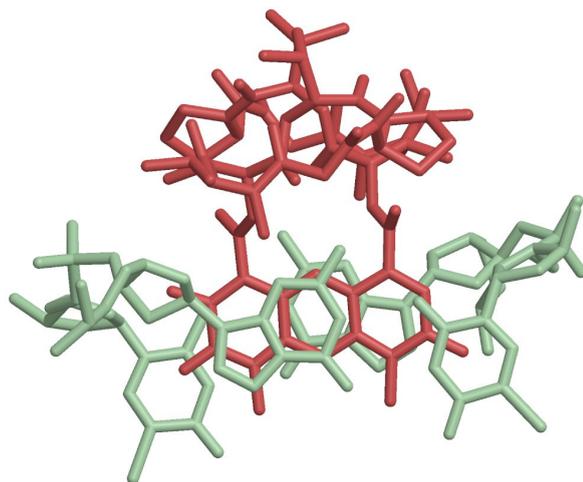


Fig. 3.29 Stacking of the phenoxazone ring of the 7-AAMD (red) between base pairs G204-BRU405 and BRU205-G404 (green).

In the case of the first complex a further π -stacking of the bromouracil ring 305 to the phenoxazone is observed (Fig. 3.23).

The typical motif of hydrogen bonding found in dsDNA complexes is present in this structure as well. The N2 of the guanines provide the hydrogen donors for the main chain O of the D-threonines of 7-AAMD. The purine residues make a total of two interactions in each complex in a reasonable distance (N2_204 \cdots O_DTH: 2.8 Å, N2_404 \cdots O_DTH: 2.8 Å, N2_304 \cdots O_DTH: 2.6 Å, N2_104 \cdots O_DTH: 3.1 Å).

3.6.4 COMPARISON WITH THE HEXAGONAL STRUCTURE

Four copies of the complex found in the asymmetric unit of the orthorhombic structure were found in the hexagonal asymmetric unit. The complexes are in general the same as in the orthorhombic case with slight differences mainly affecting the packing.

One of these differences is the absence of thymine T101 in all four models. In the orthorhombic case this base was involved in a T*A*T triplet with the symmetry related T102* and A403*. In three of the hexagonal copies the missing base is replaced either by the symmetry equivalent T302* or T401*. In the fourth copy, the triplet is formed by two bases of the same strand A403 and T401 and a symmetry related T102*. To achieve this strand **4** undergoes a high distortion in its sugar-phosphate backbone. A further consequence of this new triplet formation is the absence of the thymine interaction T302 \leftrightarrow T302* observed in the orthorhombic structure (Fig. 3.24), making T302 more flexible.

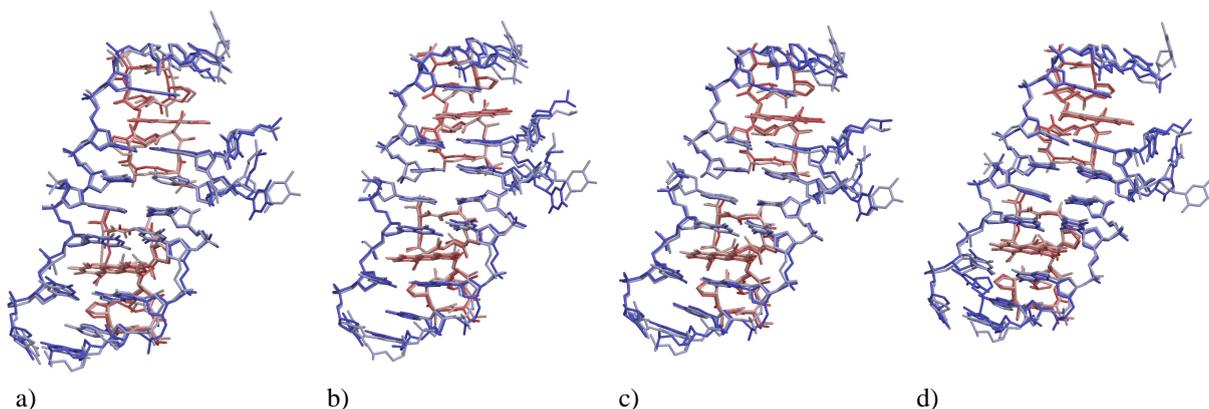


Fig. 3.30 Least squares fit of the orthorhombic copy on molecule 1 (a), 2 (b), 3 (c) and 4 (d) of the hexagonal structure. Mean rms deviations: a) 0.87 Å for 557 fitted atoms, b) 0.90 Å for 548 fitted atoms, c) 0.82 Å for 557 fitted atoms, d) 1.09 Å for 553 fitted atoms.

The least squares fit of the orthorhombic model with each of the four hexagonal copies done by the program SHELXPRO confirmed this (Fig. 3.30). The biggest deviations between the structures were observed for T302 (rms deviation of equivalent residues: T1302: 2.96 Å, T2302: 3.52 Å, T3302: 2.62 Å, T4302: 5.06 Å).

The change in the base pairing network between symmetry related strands determines the packing in the hexagonal crystals. Long tubes are located on the a,b -plane and repeat themselves in steps of $1/6 c$ with a difference in orientation of 60° (Fig. 3.31). This packing explains the puff-pastry like habit of the native crystals that could be observed as long stripes in the diffraction pattern.

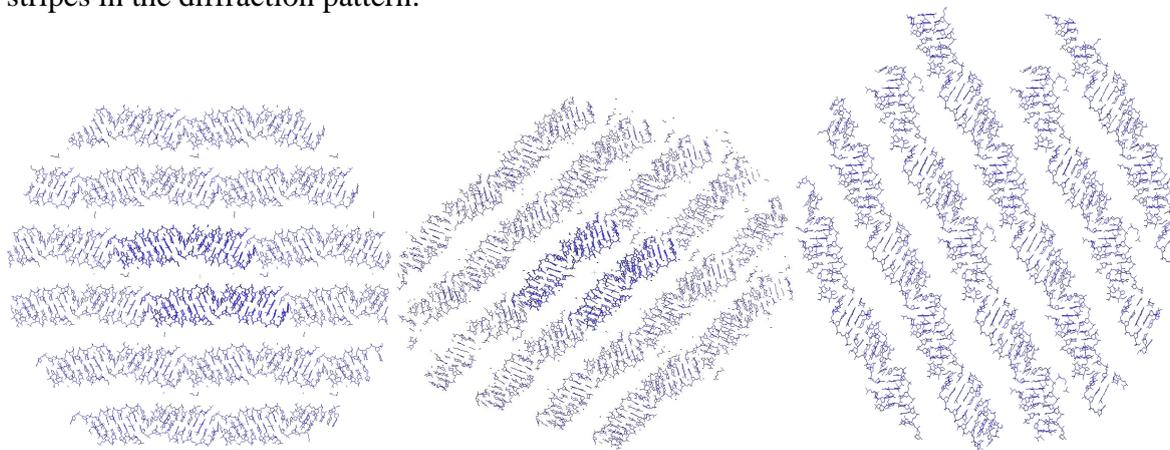


Fig. 3.31 Crystal packing in the hexagonal form.

In the orthorhombic form the packing can be described as a zigzag motif (Fig. 3.32).

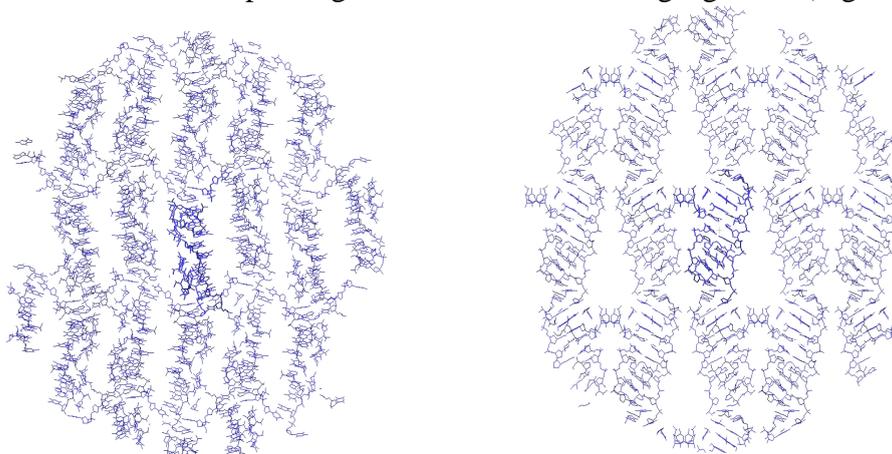


Fig. 3.32 Crystal packing in the orthorhombic case.

3.6.5 PSEUDOSYMMETRY

The orthorhombic structure could only be solved in the space group $C222_1$. Nevertheless, the “absent” reflections $h + k \neq 2n$ were measured on the synchrotron, but with a weaker intensity. The calculation of a R_{work} with only these reflections is in the same range as the actual refined R_{work} without them, proving that the reflection are not artifacts. Still, it seems that they were too weak to be observed on in house sources and get lost in the background noise.

The refinement of the structure with two of the $C222_1$ molecules in the asymmetric unit in the space group $P2_12_12_1$ yielded similar R-values to the C -centered case ($R_{\text{work}} = 0.24$, $R_{\text{free}} = 0.31$). Only small differences are found between the two space groups. The fitting of both copies of $P2_12_12_1$ on the $C222_1$ structure with SHELXPRO resulted in rms deviations of 0.3864 Å and 0.3823 Å respectively. The two fits are shown in Fig. 3.33.

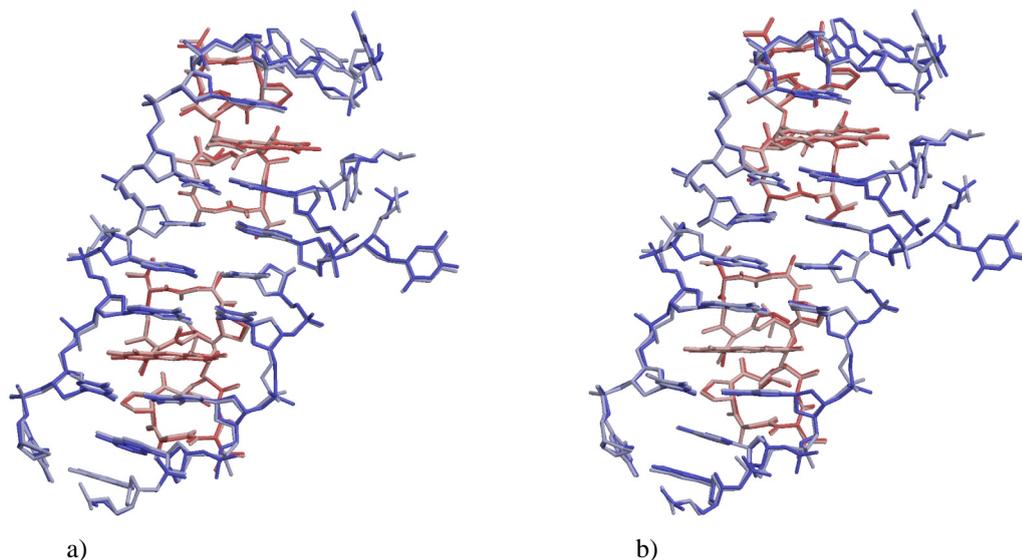


Fig. 3.33 a) Fit of the $C222_1$ model on the first copy of the $P2_12_12_1$ structure, b) Fit of the $C222_1$ model on the second copy of the $P2_12_12_1$ structure.

Deviations between atom positions greater than two sigma are observed for the residues T101, G204, BRU205, T302, T402, A403, G404, BRU405 and are located primarily in the phosphate backbone or in bases not involved in stacking. The pyrimidines or purines involved in pairing show a very good overlap of their atoms.

Nevertheless there must be some difference between the structures in $C222_1$ and $P2_12_12_1$ causing this symmetry dilemma. The “absent” reflections contain structural information, which cannot be revealed unless the structure is solved in $P2_12_12_1$.

THE D,L-ALTERNATING PEPTIDE

H-(L-TYR-D-TYR)₄-LYS-OH

4.1 INTRODUCTION

Peptides and polypeptides with a regular sequence of enantiomeric residues (L and D) along the chain are well known in nature. The most distinguished example is the pentadecapeptide Gramicidin A, an antibiotic with alternating L- and D- residues along the peptide sequence. Gramicidin A is a naturally occurring product of *Bacillus brevis* known to form ion channels in synthetic and natural membranes. It seems to be selective for monovalent cations such as H⁺, Tl⁺, NH₄⁺ and the alkali metals. Several crystal structures of gramicidin have been reported, all of them being double stranded antiparallel β -helices either complexed with cations or in an uncomplexed form with the side chains oriented towards the membrane. The number of amino acids per helical turn differs from structure to structure, defining several sizes for the channel to accommodate various unsolvated ions, from K⁺ to Cs⁺, and showing a certain flexibility of the helix.

Structures for further D,L-alternating peptides were obtained for oligomers like *t*-Boc-(L-Val-D-Val)₄-OMe and *t*-Boc-(L-Phe-D-Phe)₄-OMe already by Lorenzi et al. (1989) showing a common motif of two antiparallel helical strands interacting by backbone hydrogen bonding. In all cases the dimer formed can be regarded as a cylinder with a hydrophilic inner core formed by the peptide backbone units and a hydrophobic exterior of the side chain groups.

All crystals of the known oligomers were grown from solutions containing organic solvents like chloroform, methanol or ethanol. It was interesting to study the structure of D,L-alternating peptides in water in order to analyze their conformational behavior in an aqueous medium. UV-absorption investigations in the group of Prof. Diederichsen, Institute of Organic Chemistry Göttingen, of H-(L-Phe-D-Phe)₄-OH in water implied a destacking of the side chain chromophores by sigmoidal increase of the absorption with raising temperature, which indicated a quite surprising stability of T_m = 34 °C.

In collaboration with the group of Prof. Diederichsen, Institute of Organic Chemistry, University of Göttingen, the crystallization of the peptide H-(L-Tyr-D-Tyr)₄-Lys-OH from water and its structural characterization were undertaken.

4.2 CRYSTALLIZATION

The D,L-peptide was synthesized by Andrea Küsel with the Fmoc-solid phase method (Chan, White) and purified via HPLC. Native crystals were grown by vapor diffusion at 20° C in hanging drops (4 μL) from a peptide solution with a concentration of 10 mg/ml using 30% MPD, 0.02M sodium bromide and 0.5M ammonium sulfate in 100 mM HEPES buffer (pH 7.5) as a precipitant in a 1:1 ratio. Single crystals grew over a period of one week (Fig.4.1).

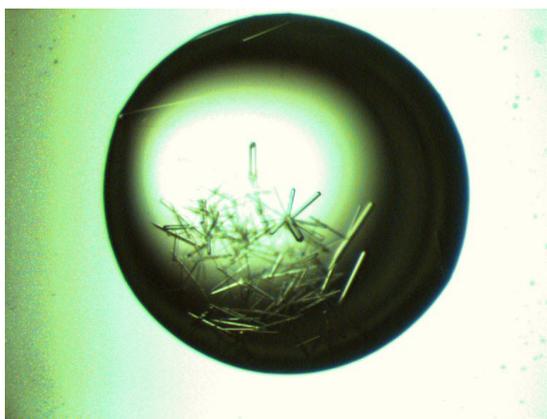


Fig. 4.1 Crystals of H-(L-Tyr-D-Tyr)₄-Lys-OH.

These crystals were extremely sensitive and dissolved immediately upon removal from the drops. Their reproduction at 4° C from the same solutions solved this problem.

Crystals to be used for iodide soaking were grown under the same conditions but in the absence of sodium bromide. These crystals were soaked for one minute in the precipitant solution containing 1M lithium iodide. In both cases no cryoprotectant was required because the crystallization solution had a glass-like freezing behavior. Both the native and the soaked crystals belongs to the tetragonal space group $P4_32_12$, with cell dimensions $a = b = 27.99$ (3) Å, $c = 78.93$ (8) Å, $V = 61878.27$ Å³. The asymmetric unit contains two peptide molecules and 40% solvent.

4.3 DATA COLLECTION

Reflection data statistics for all datasets are shown in Table 4.1.

Table 4.1: Data collection statistics						
	Native crystal			Iodide-soaked crystal		
	1.5418 Å in-house 1	0.8015 Å X13 2	1 + 2 Merged	1.5418 Å in-house 3	0.9540 Å BW7A 4	3 + 4 Merged
crystallographic data						
<i>a</i> (Å)	28.165	27.996		28.007	28.019	
<i>c</i> (Å)	79.317	78.910		79.368	79.388	
space group	<i>P</i> ₄ ₃ ₂ ₁					
Resolution (Å)	2.20	1.20	1.30	2.44	2.49	2.44
Uniq. reflections	22183	10531	8341	6624	17176	1332
$\langle I/\sigma(I) \rangle$	10.33 (7.54)	3.95 (1.50)	14.35 (1.50)	6.30 (5.14)	6.92 (6.71)	27.08 (8.05)
R_{merge}	0.055 (0.211)	0.071 (1.285)	0.089	0.137 (0.117)	0.085 (0.331)	0.106
completeness (%)	96.7 (80.7)	98.7 (96.9)	99.0 (96.9)	77.5 (20.6)	99.8 (98.9)	94.0 (58.3)
redundancy	11.6 (3.51)	8.00 (6.22)	-	4.71 (0.41)	8.02 (7.36)	-

Table 4.1 Data collection statistics. Values in brackets refer to the outer 0.1 Å wide resolution shell.

4.3.1 NATIVE DATA

Two low temperature native datasets were collected. The first measurement was performed in house on a Bruker SMART 6000 CCD detector using a crystal of the size 0.2 x 0.2 x 0.1 mm³ mounted in a loop at a wavelength of 1.5418 Å and a temperature of 100K. A thin slicing mode was used with an oscillation angle of 0.2°, a crystal-to-detector distance of 550 mm and an exposure time of 60 sec. A total of 22183 data to 2.2 Å resolution were measured. Data integration and scaling was performed using the programs PROTEUM and SADABS (Bruker AXS, 2000).

The crystal was then frozen and data were collected at the EMBL, c/o DESY, Hamburg, on the X13 synchrotron beamline with an X-ray wavelength of 0.8015 Å using a MAR CCD detector. A high resolution dataset to 1.2 Å with an oscillation angle of 0.5° and a low resolution dataset to 2.45 Å also with an oscillation angle of 0.5° were measured. Due to fluctuations in the detector temperature two separate high resolution datasets had to be integrated with DENZO (Otwinowski & Minor, 1997), but could be merged together with

the low resolution data during scaling with SCALEPACK (Otwinowski & Minor, 1997). This resulted in a lower quality of the synchrotron data in comparison to the in house data in spite of the better resolution.

4.3.2 IODIDE-DERIVATIVE DATA

From the iodide-soaked crystal a first dataset was measured in house on a mar345 image plate with Bruker rotating Cu-anode and osmic mirror system ($\lambda = 1.5418 \text{ \AA}$). Data were collected to a resolution of 2.44 \AA at temperature of 100K. The dataset was originally intended as a test for the quality of the crystal, but was nevertheless used for the structure solution later on. Therefore the completeness is rather low (77.5%). The processing was performed with DENZO and SCALEPACK.

The frozen crystal was transported to the EMBL, c/o DESY, Hamburg, and data were collected on synchrotron beamline BW7A at 100K with an X-ray wavelength of 0.954 \AA using a mar CCD unsupported-m165 detector. Surprisingly, the resolution attained is slightly lower than in house (2.49 \AA), which could be a result of crystal damage during the freezing. These data were processed with HKL2000.

4.4 STRUCTURE SOLUTION

The two native datasets were collected from the same crystal and could be merged into a single dataset with 10560 averaged reflections. The R_{merg} is 0.09 and shows a relatively good fit of the common reflections. The merged native dataset was truncated at 1.3 \AA . As the crystal was grown in a solution containing bromine ions, the data was checked for anomalous signal, but none could be detected (Fig. 4.2).

```
Anomalous signal/noise ratios (1.0 is random). The first line is based on
input sigmas, the second on variances of F+ and F- (if not already averaged):
Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.5 - 2.2 - 2.0 - 1.8 - 1.6 - 1.4 A

0.93 1.13 0.68 0.57 0.82 0.89 1.15 1.31 1.29 1.17 1.03 1.02

91.7 Neighbors used on average for F+/F- local scaling
Rint(anom) = 0.0584 before and 0.0592 after local scaling
```

Fig. 4.2 Anomalous signal/noise ratio for the native dataset as calculated from XPREP.

4.4.1 DIRECT METHODS

The molecule consists of 115 non-hydrogen atoms and the resolution is quite high, so it was tried to solve the structure with SHELXD using *ab Initio* methods. It is usually more efficient to search for ca. 25% less than the total number of non-solvent atoms, so a target of 180 peaks was defined. During the peaklist optimization more than the expected number of atoms were specified because this procedure involves the elimination of the wrong ones. The minimum *E*-value was set to 1.5. In a total of 50 tries SHELXD produced correlation coefficients in the region of 30-35%, even after variation of the target peak number and the *E*-value, indicating that no solution was found. In a further attempt the main-chain atoms of gramicidin A (78 atoms) were used as starting atoms for an expansion of the structure. SHELXD resulted a correlation coefficient of 37.93%. The most plausible explanation for the failure of *ab initio* methods in the case of such a small structure at this rather high resolution is the bad quality due to detector instability of the Hamburg data set, which alone contributes to the high resolution shells of the merged native data.

4.4.2 SAD

The iodide derivative datasets had also been collected on the same crystal and were merged with an R_{merg} of 0.11. The merged dataset contains 2139 averaged reflections at 2.44 Å resolution. The anomalous signal to noise ratio of the datasets is shown in Fig. 4.3. In general the iodide anomalous signal is rather weak for all three sets and becomes insignificant for a resolution higher than 3.8 Å. Thus, it was not possible to solve the structure with SAD.

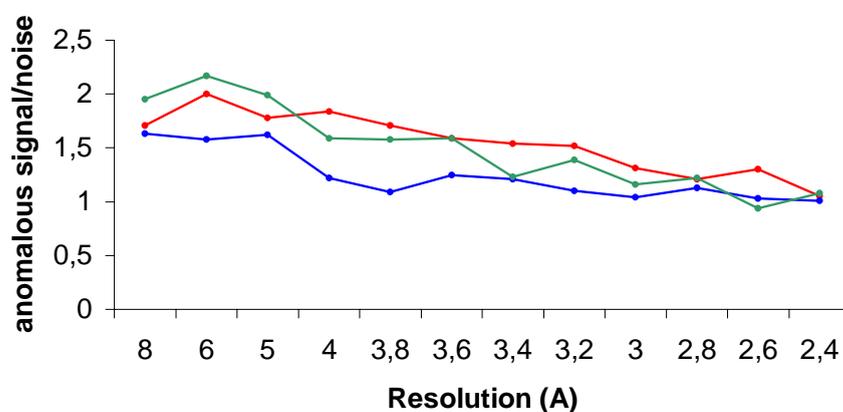


Fig. 4.3 Anomalous signal to noise ratio against resolution for data set 1 (red), for data set 2 (blue) and for the merged data set (green).

4.4.3 SIRAS

After local scaling of the merged native and the merged derivative data there resulted a R_{merg} of 0.17. A R_{merg} value in that region (20-25%) indicates a good similarity between the data sets, making the calculation of SIRAS differences possible, but is nevertheless high enough to suggest that some change took place during the soaking of the crystal. The resolution cutoff of the SIRAS data is of great importance for the success of the substructure solution, especially in case of a weak anomalous signal. By excluding high resolution shells with none or very low anomalous signal the overall anomalous signal to noise ratio is improved. In this case the data was truncated at 2.5\AA (Table 4.1). In a run of 1000 tries with a minimum E -value of 1.3 SHELXD found six heavy atom sites, two of which were strongly occupied and two very weakly. 400 cycles of density modification with 40% solvent content were performed with SHELXE for both enantiomorphs. The mean estimated map CC was calculated for both cases and was highest for $P4_32_12$ (Fig. 4.4) identifying it as the correct space group.

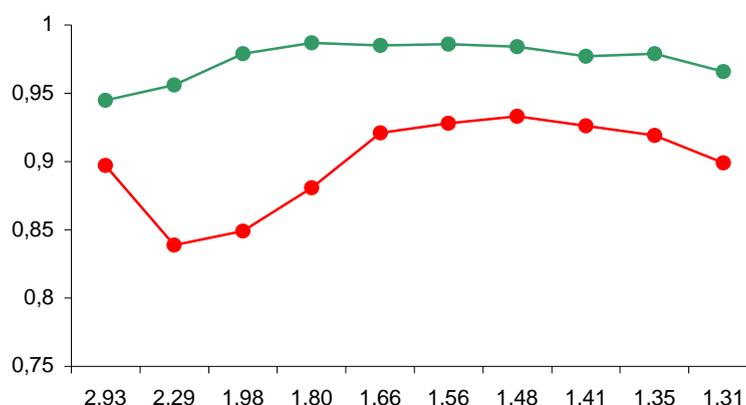


Fig. 4.4 Mean estimated map CC against resolution (\AA) from SHELXE after 400 runs for space groups $P4_32_12$ (green) and $P4_12_12$ (red).

For this structure both the contrast and the connectivity showed significant differences starting from the two heavy atom enantiomorphs, being for $P4_12_12$ somewhat lower ($\text{contrast} = 0.373$, $\text{connectivity} = 0.877$) than for $P4_32_12$ ($\text{contrast} = 0.551$, $\text{connectivity} = 0.902$). Last, a *pseudo-free correlation coefficient* was calculated leaving out 10% of reflections at random and doing one cycle of density modification ($P4_32_12$: 84.89%, $P4_12_12$: 69.15%). The electron density map obtained could be traced by hand using the structure of a known d,l-alternating peptide (t -Boc-(L-Phe-D-Phe) $_4$ -OMe) (Lorenzi, 1989) as a first model.

4.5 STRUCTURE REFINEMENT

The model was subject to least-squares refinement with SHELXL97 against F^2 of the synchrotron data set using a conjugate gradient algorithm. A set of reflections representing 5% of the overall reflections was chosen in thin shells, in case non-crystallographic symmetry should be applied, to define an R_{free} set for cross-validation. The refinement stages are summarized in Table 4.2.

Action	Peptide atoms	Water	Parameters	Restraints	R1 [$I > 2\sigma(I)$ / all]	Rfree [$I > 2\sigma(I)$ / all]
model building 1.3 Å	209	-	843	903	0.3145/0.3432	0.3590/0.3810
water addition 1.3 Å	209	16	907	903	0.2664/0.2943	0.2920/0.3130
model completion	211	16	915	908	0.2528/0.2799	0.2787/0.2985
resolution cutoff 1.4 Å	211	16	915	908	0.2464/0.2650	0.2783/0.2930
anisotropic scaling 1.3 Å	211	16	927	908	0.2021/0.2265	0.2314/0.2483
anisotropic refinement 1.3 Å	211	16	1975	2570	0.1836/0.2059	0.2346/0.2479
water addition 1.3 Å	211	19	939	907	0.1971/0.2216	0.2332/0.2493
hydrogen addition 1.3 Å	364	19	1977	2574	0.2042/0.2285	0.2372/0.2537
Final	211	19	939	1	0.1989/0.2230	-

Table 4.2 Refinement progress.

The refinement stages are summarized in Table 4.2. In the first cycle the double helix was built by tracing the map and using the d,l-alternating peptide (*t*-Boc-(L-Phe-D-Phe)₄-OMe) as a template. Geometric restraints for 1,2- and 1,3-distances were defined using the Engh & Huber (1991) values as targets. Furthermore, planarity, anti-bumping and chiral volume restraints were defined. The diffuse solvent region was accounted for by a two parameter model after Babinet's principle (Moews & Kretsinger, 1975). There resulted an initial R_{work} of 0.31 ($R_{\text{free}} = 0.36$). As a next step water molecules were added manually to the map, selecting among the highest electron-density peaks those making reasonable contacts. The approximately spherical shape of the $F_o - F_c$ electron density corresponding to water molecules was confirmed in the σ -weighted map. The model was completed to a R_{work} of 0.25 and $R_{\text{free}} = 0.28$. Truncation of the data to a resolution of 1.4 Å improved the R_{work} , but did not affect the R_{free} significantly and was therefore abandoned.

Taking into account the detector temperature fluctuations and the long needle-like shape of the crystals an anisotropic scaling correction was considered. In this procedure twelve anisotropic scaling parameters are refined for the otherwise isotropic structure (Usón et al, 1999). Nevertheless, an anisotropic refinement of the atomic B values, which is recommended at this resolution, led to better results for the R_{work} than the scaling correction (see Table 4.2). One should bear in mind though that anisotropic scaling involves only 12 extra parameters leading to a data to parameter ratio of 8.5:1, whereas the ratio in the case of the anisotropic refinement is 4:1. This fact is also reflected by the minimal improvement of the R_{free} values (less than 0.01%), indicating that the effect on the R_{work} is purely cosmetic and that the anisotropic scaling should be favored.

The inclusion of hydrogen atoms at geometrically calculated positions and their refinement using a riding model did not lead to any improvement, but increased the R-values. Therefore the H-atoms were removed. As a further step the structure was refined to convergence with 19 water molecules to the final values $R_{\text{work}} = 0.20$ and $R_{\text{free}} = 0.23$. Finally, full-matrix least-squares refinement was performed against all data (including the R_{free} set) to calculate standard uncertainties (s.u.). The refinement proved to be unstable so the parameters had to be kept at fixed values to prevent excessive shifts.

Non-crystallographic constraints were not applied in the model. More details of the refinement statistics are shown in Table 4.3.

Unit cell parameters (Å)	$a = b = 27.99(3), c = 78.93(8)$
Space group	$P4_32_12$
No. of unique reflections	8289 (7883 without R_{free})
Resolution range	20-1.3
R_{factor} ($I > 4\sigma$)	0.1971
R_{free} ($I > 4\sigma$)	0.2332
No. of water molecules	19
Data/restraints/parameter	7883/907/939
R.m.s. deviations from idealized geometry	
Bond lengths (Å)	0.011
1,3-distances (Å)	0.028
Zero chiral volumes (Å ³)	0.072
Non zero chiral volumes (Å ³)	0.100
Distances from restraint planes (Å)	0.179
Mean B factor (Å ²)	
Main-chain atoms	10.31
Side-chain atoms	18.06
Solvent atoms	24.38

Table 4.3 Refinement statistics.

4.6 STRUCTURE OF THE H-(L-TYR-D-TYR)₄-LYS-OH

The asymmetric unit of the tetragonal H-(L-Tyr-D-Tyr)₄-Lys-OH contains a single dimer and 19 water molecules in the asymmetric unit. A stereo view of the dimer is shown in Fig. 4.5.

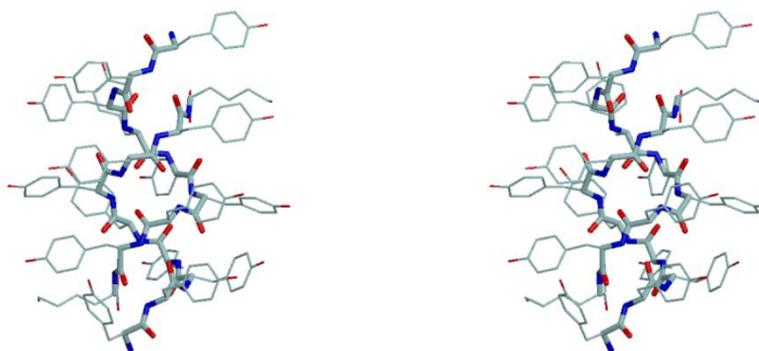


Fig. 4.5 Stereo projection of the right-handed antiparallel double-stranded β -helix of H-(L-Tyr-D-Tyr)₄-Lys-OH.

It is a right-handed double-stranded antiparallel $\beta^{5,6}$ helix with an overall length of 19.7 Å. The average inner channel diameter as calculated for the equivalent C α atoms of each strand is ca. 6 Å, varying from 4.7 Å to 7.2 Å. Taking van der Waals radii into account there results an inner radius of 2.2 Å, a value which allows the transfer of alkali metal ions through the channel, but not water molecules (Fig. 4.6).

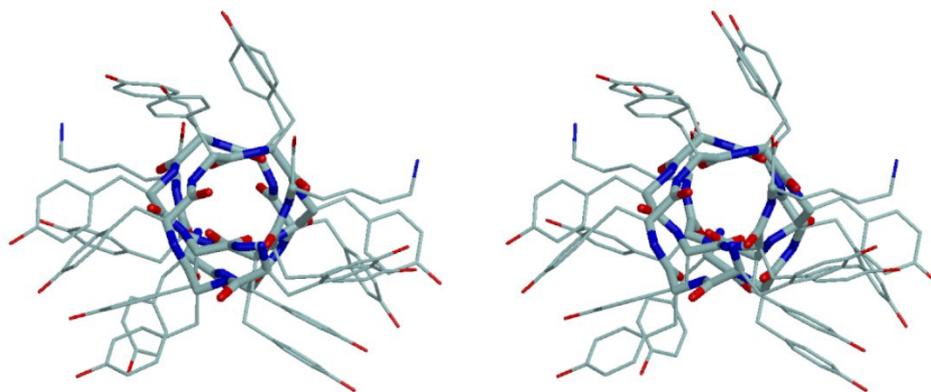


Fig. 4.6 Stereo projection of H-(L-Tyr-D-Tyr)₄-Lys-OH along the double helix axis.

As is typical for alternating d,l-peptides and also for β -sheets in general the association of the two helical chains is stabilized by systematic NH \cdots O=C hydrogen bonds between donor and acceptor groups of different strands. Seven N-H groups of one chain are H-bonded to seven O=C groups of the other and vice versa (average N \cdots O distance 2.9 Å). The two remaining N-H groups of each chain are connecting the dimer in either direction of the helix with the next symmetry equivalent dimer. In that way pipes are formed along the crystallographic z axis.

These pipes are linked to each other via a complicated water network. Three tyrosines of one chain and five of the other are connected with symmetry equivalents or with each other via 17 water molecules. The H-bonds are formed by the OH-groups of the tyrosines, which either interact over the water molecules or directly with the OH-groups of symmetry equivalent residues (Fig. 4.7).

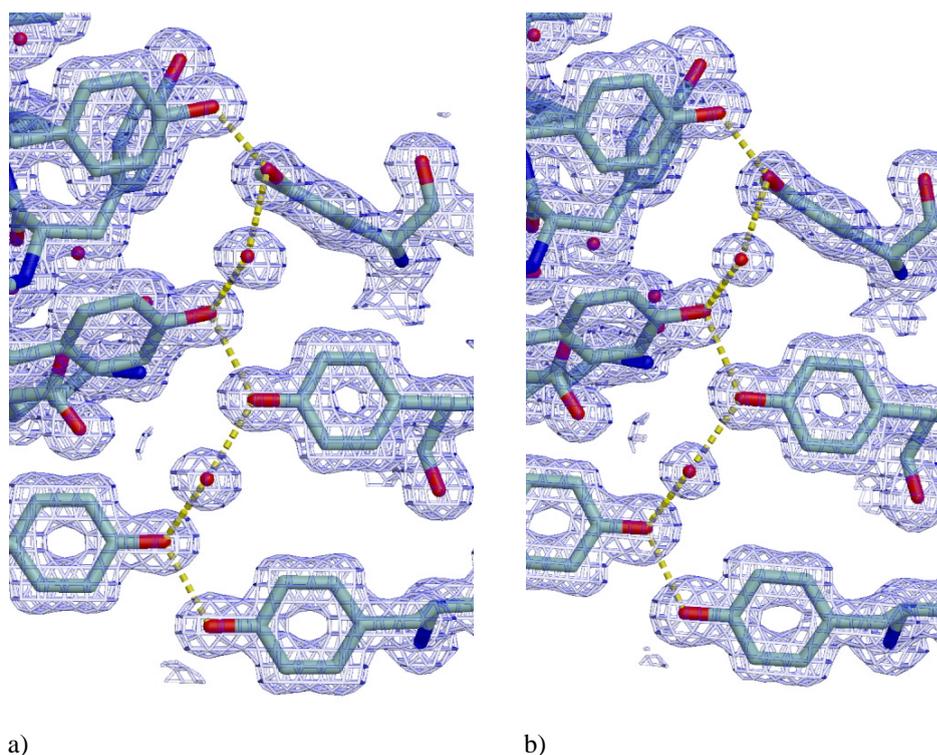


Fig. 4.7 a) View of the electron density of part of the water network between two symmetry related molecules. The hydrogen bonds are shown as yellow dashed lines, water oxygen atoms are red. The sigmaA weighted map was obtained after the last refinement of the structure with SHELXL, b) View of the electron density of part of the water network between two symmetry related molecules as seen in the first experimental map obtained with SHELXE. Again hydrogen bonds are shown as yellow dashed lines, while water oxygen atoms are red.

Due to this tight water network the OH-groups of the tyrosines are clustering in the direction of the contact surface with the next chain (Fig.4.8).

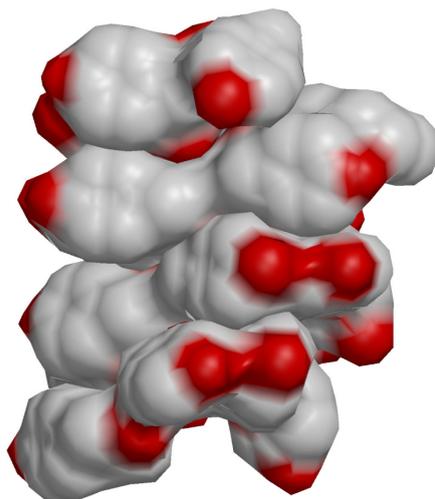


Fig.4.8 Surface of the helix of H-(L-Tyr-D-Tyr)₄-Lys-OH, as calculated with msms. Oxygen atoms are shown in red.

In this way a π -stacking of the aromatic rings is prohibited. This stacking has nevertheless been observed for the H-(L-Phe-D-Phe)₄-OH peptide, whose structure was determined by Lorenzi et al. (1989). The reason for this difference is probably are the water molecules interacting with the tyrosines. The crystals of the phenylalanine substituted peptide were grown in organic solvents. An overview of the conformational parameters of the helix is shown in Table 4.4.

Angle	L	D	L	D	L	D	L	D
ϕ (N-C α)		150	-100	150	-111	153	-114	147
ψ (C α -C)	119	-86	150	-130	140	-74	143	-111
χ_1 (C α -C β)	175	168	-64	-177	171	158	-60	176
χ_2 (C β -C γ)	83	104	92	113	71	41	-77	-79

a) Conformational parameters of H-(L-Tyr-D-Tyr)₄-Lys-OH, strand 1.

Angle	L	D	L	D	L	D	L	D
ϕ (N-C α)		151	-94	136	-106	149	-116	145
ψ (C α -C)	117	-89	158	-140	146	-72	146	-116
χ_1 (C α -C β)	-63	165	-68	75	-65	168	-60	-180
χ_2 (C β -C γ)	97	-119	102	-111	-91	-4	105	-81

b) Conformational parameters of H-(L-Tyr-D-Tyr)₄-Lys-OH, strand 2.

Table 4.4 Conformational parameters of the two strands.

The average ϕ / ψ torsion angles are $147.33^\circ \pm 1.55^\circ$ / $-102.25^\circ \pm 1.68^\circ$ and $-119.05^\circ \pm 1.72^\circ$ / $139.91^\circ \pm 1.53^\circ$ for the D- and L-residues respectively. The side chain conformation of the tyrosines is defined primarily by the χ_1 and the χ_2 torsion angles. For χ_1 , in one helical chain there are four g^- , three t and one g^+ conformations, whereas in the other the preferred conformation is the t (six times) with two g^- . This observation is slightly deviating from the average geometric parameters reported in the literature (Benedetti, 1983), which show a preference for the g^- conformation, and is probably an effect of the H-bonds to the symmetry equivalent chains. The distribution of the χ_2 torsion angles can be expressed in the range from 0° to 180° , because variation of χ_2 has very nearly twofold symmetry in the tyrosine

residue. With the exception of the sixth D-tyrosine residue the values are in the expected range of 70° to 120° .

The least squares fit of the two strands was done by the program SHELXPRO (Fig. 4.9). The biggest deviations between the structures were observed for the tyrosine side chains (rms deviation of equivalent residues: Tyr1 3.77 Å, Dty2 1.52 Å, Tyr3 1.11 Å, Dty4 3.10 Å, Tyr5 3.16 Å, Dty6 2.10 Å, Tyr7 1.97 Å, Dty8 1.01 Å, Lys9 1.15 Å).

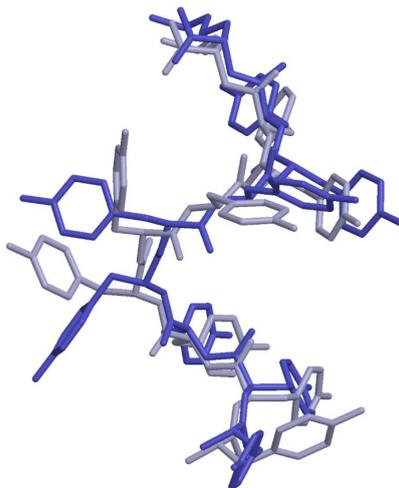


Fig. 4.9 Least squares fit of the two strands.

MOLECULAR MODELING STUDIES ON hER α LBD WITH ORGANIC LIGANDS

5.1 INTRODUCTION

5.1.1 AIM OF THE PROJECT

Regulation of gene expression in a chemical dependent way can be achieved by fusion of a transcription factor of interest to a receptor ligand binding domain that can be activated by a specific ligand. This has been previously demonstrated in plants using different inducible systems. A well-known system is the fusion of factors that regulate trichome development (Lloyd *et al.*, 1994), leaf morphology (Aoyama *et al.*, 1995) and flowering time (Simon *et al.*, 1996) with the ligand binding domain of the glucocorticoid receptor (GR). A more recent regulatory construct is the ecdysone agonist inducible transcription, applied in transgenic tobacco plants (Martinez *et al.*, 1999). This system is composed of the *Heliothis virescens* ecdysone receptor, which belongs to the steroid/retinoic acid/thyroid receptor superfamily, and the transactivation domain of the *Herpes simplex* VP16 protein fused to the DNA binding and transactivation domain of a mammalian glucocorticoid receptor. A commercially available non-steroidal ecdysone agonist, RH5992 (tebufenozide), is used as an inducer.

The development of a novel plant gene expression regulatory system with a mutant of the ligand binding domain of the human estrogen receptor α (hER α) as the chemically inducible species is the aim of a collaboration within the groups of Prof. Sheldrick, Department of Structural Chemistry, Prof. Tietze, Institute of Organic Chemistry, and Prof. Gatz, Plant Institute, all in the University of Göttingen. In this project, the natural ligand of hER α , the steroid estradiol, will be replaced by synthetic environment-friendly organic ligands designed to activate a mutated receptor LBD.

This thesis describes the molecular modeling studies of hER α LBD and their mutants complexed with organic compounds.

5.1.2 THE HUMAN ESTROGEN RECEPTOR ALPHA LIGAND-BINDING DOMAIN

5.1.2.1 hER α

The human estrogen receptor (hER) is a transcription regulator that belongs to the superfamily of nuclear receptors. The nuclear receptors (NR) fall into two groups, the steroidhormone receptors (glucocorticoid GR, progesterone PR, mineralocorticoid MR, androgen AR and estrogen ER receptors), which act as homodimers, and the type II subfamily (thyroid TR, retinoic acid RAR, vitamin D VDR, peroxisome proliferator-activated PPAR and orphan receptors), which usually heterodimerize with the retinoid X receptor (RXR) to achieve stable DNA binding and transcriptional activity.

The structural domains of the estrogen receptor, which also represent the general structure of nuclear receptors, are shown in Fig 5.1.

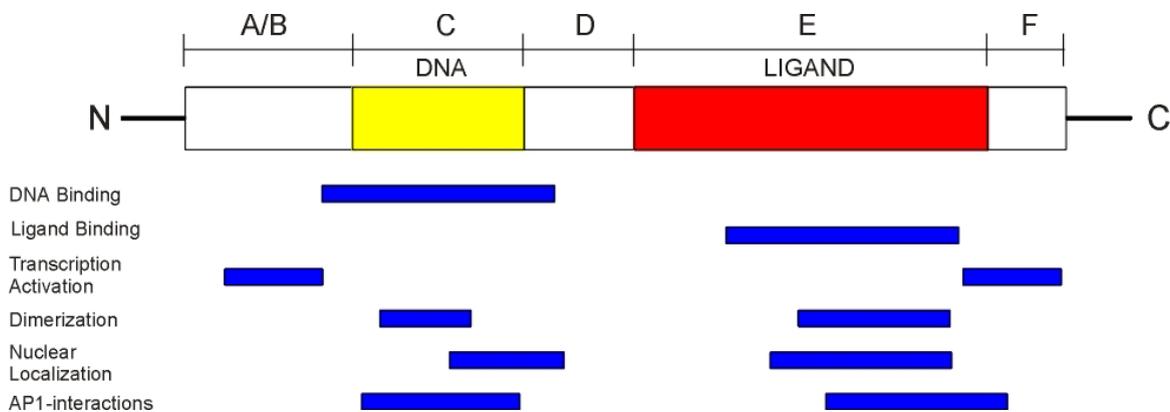


Fig. 5.1 Structural domains of the ER (Jones & Petkovitch, 1996).

The A/B domain is located at the N-terminus and contains a ligand-independent transactivation domain, as well as sites for interaction with other proteins. The DNA binding domain C is highly conserved in sequence and contains two zinc fingers involved in DNA binding, but also regions responsible for homo- and heterodimerization and nuclear localization. The D domain is the so-called hinge region and seems to accommodate dimerization. The E domain is the ligand binding domain (LBD) and highly conserved

among receptors of the same family sharing the same ligand. The function of the F domain is unknown.

The estrogen receptor action is shown in Fig. 5.2.

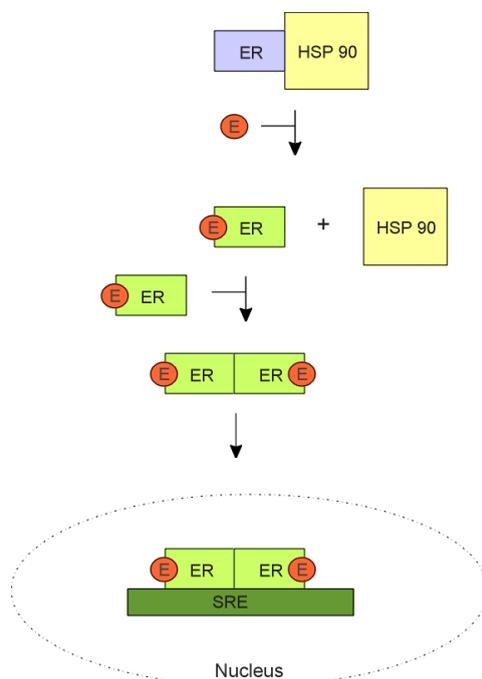


Fig. 5.2 Mechanism of estrogen receptor function. ER = estrogen receptor, HSP 90 = heat shock protein 90, E = estradiol, SRE = steroid response element.

In the absence of ligand, the estrogen receptor is associated with cytoplasmic heat-shock protein 90 (HSP 90). The binding of estradiol results in an activating transformation event that displaces the HSP and facilitates the dimerization of the receptor. As a homodimer ER then translocates to the nucleus and binds to steroid response elements (SRE) in the promoters of target genes. The complex formed is involved in increasing the efficiency of assembly of the transcription initiation complex. However, as RNA polymerase II does not interact directly with the estrogen receptor, additional factors are used to allow the two proteins to communicate. One class of factors are the coactivator proteins, that interact only with the agonist-activated receptor. Members of the p160 family of coactivators recognize the active NR LBD through a short signature sequence motif LXXLL (where L is leucine and X is any amino acid), known as the NR box.

Estrogen receptors have been investigated for a variety of mammalian systems. The current thesis is focusing only on the human ER. Its importance as a therapeutic target lies on one hand in the fact that it affects the expression of specific genes, many of which have been well studied, and on the other hand in its direct regulation by small molecules, making it easier to modify its transcriptional activity. A number of human diseases occur as a result of

abnormal production of hormones. Osteoporosis for example is common for post-menopausal women in which bone mass is significantly reduced as a result of decreased production of estrogen. Furthermore, the progression of breast, ovarian, and endometrial cancers can be related to an abnormal response to estrogen.

Two subtypes of the human estrogen receptor are known, the α and β subtype. ER α and ER β share significant amino acid homology in their ligand-binding domains (~55%) with only two changes at L384(ER α)/M336(ER β) and M421(ER α)/I373(ER β) falling into the binding pocket. Both isoforms bind estradiol with equivalent affinity. However, they are not pharmacologically equivalent as they show significant differences in ligand preferences. Tetrahydrochrysen (THC) for example acts as an agonist for ER α and as an antagonist for ER β .

5.1.2.2 EXISTING X-RAY STRUCTURES

While no structure of the DNA-binding domain of hER has been yet determined, there are nine x-ray structures of the LBD of the α subtype and three of the β subtype deposited in the Protein Data Bank (Table 5.1).

Species	PDB ID	Resolution (Å)	Ligand	Conformation	Literature
hER α	3ERD	2.00	DES + GRIP 1 NR Box II peptide	agonistic	Greene <i>et al.</i>
hER α	1G50	2.90	Estradiol	agonistic	Eiler <i>et al.</i>
hER α	1ERE	3.10	Estradiol	agonistic	Brzozowski <i>et al.</i>
hER α	1GWR	2.40	estradiol + TIF 2 NR Box III peptide	agonistic	Pike <i>et al.</i>
hER α	1QKU	3.20	Estradiol	agonistic	Moras <i>et al.</i>
hER α	1L2I	1.95	THC + GRIP 1 NR Box II peptide	agonistic	Shiau <i>et al.</i>
hER α	1UOM	2.28	tetrahydroiso- quinoline derivative	agonistic	Stark <i>et al.</i>
hER α	1A52	2.80	Estradiol	antagonistic	Tanenbaum <i>et al.</i>
hER α mutant C381S, C417S, C530S	1QKT	2.20	Estradiol	antagonistic	Gangloff <i>et al.</i>
hER α	1ERR	2.6	Raloxifene	antagonistic	Brzozowski <i>et al.</i>
hER β	1QKM	1.80	Genistein	antagonistic	Pike <i>et al.</i>
hER β	1QKM	2.25	Raloxifene	antagonistic	Pike <i>et al.</i>
hER β	1NDE	3.00	Triazin Modulator	antagonistic	Lambert <i>et al.</i>
hER β	1L2J	2.95	THC	antagonistic	Shiau <i>et al.</i>

Table 5.1 Existing X-ray structures of hER in the Protein Data Bank.

Both subtype receptors show different conformations. Six structures of hER α show an agonistic conformation, whereas in three cases (1QKT, 1A52, 1ERR) the LBD has an antagonistic form. Although the ligand for 1A52 and 1QKT is estradiol, the conformation is antagonistic due to crystal packing effects in the first case and due to triple mutation (C381S/C417S/C530S) in the second case. In 1ERR raloxifene acts as an antagonist.

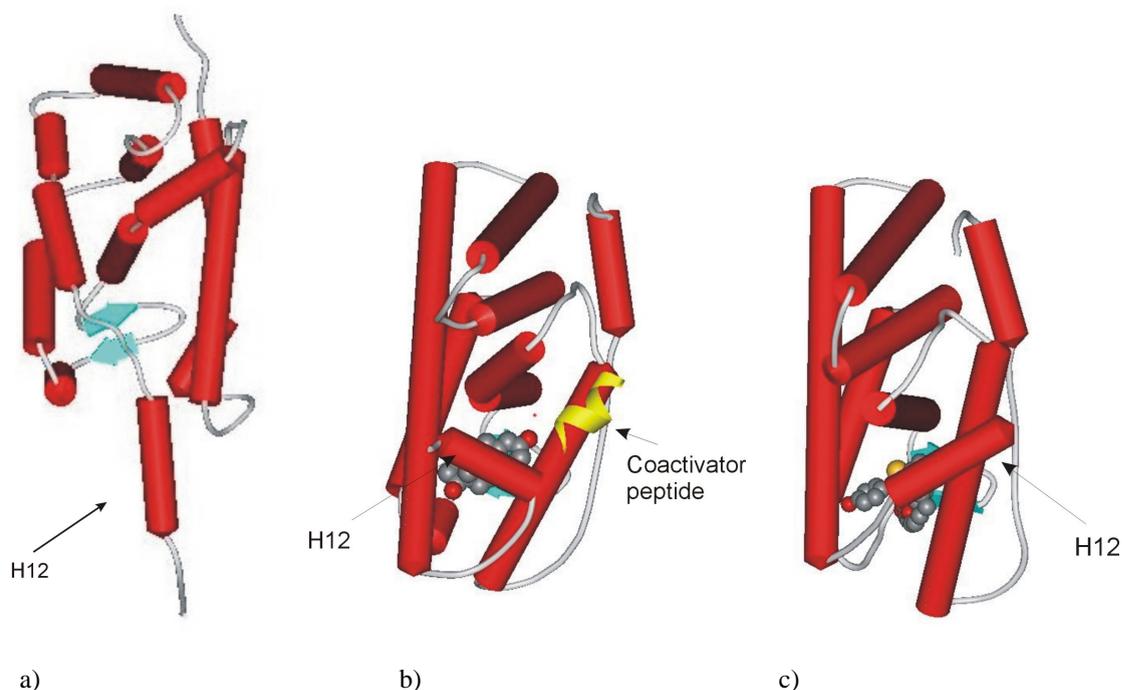


Fig. 5.3 a) The apo-LBD from the crystal structure of the apo-retinoic acid receptor (RAR) α LBD (PDB ID: 1LBD), b) Agonistic conformation of hER α LBD with estradiol, c) Antagonistic conformation of hER β LBD with raloxifene.

In both conformations (Fig. 5.3 b + c) the protein of the α subtype consists of 12 α -helices and one small β -sheet. The secondary structure elements are arranged in a so-called α -helical sandwich, which is common for all NR LBDs. The hormone-binding cavity lies immediately below H6 and is sandwiched between the outer flanking layers. Agonist binding induces a conformational rearrangement in the LBD resulting in the creation of a specific binding site for the NR-box module of nuclear coactivators. The helix H12 serves as a lid closing the entrance to the ligand binding pocket. The binding of an antagonist forces helix H12 to reorient and take the position of the NR box module, thus blocking co-activator binding. The structure of the apo-form of hER is not known, but the crystal structure of apo-RAR LBD has been determined (Moras et al, 1995) and is shown in Fig. 5.3 a. Here H12 is turned away from the rest of the protein. It is clear that the binding of a ligand causes conformational changes, suggesting that the interactions with helix 12 are important for the further agonistic/antagonistic properties of the receptor.

However, there is a third group in the classification of hER ligands: the partial agonists, known as SERMs (selective estrogen receptor modulators). These molecules have the ability to act as both agonists and antagonists depending on the tissue context as well as the ER isoform. The pharmacological effect of SERMs make them important pharmaceutical targets for hormone replacement in menopausal women and for chemotherapeutic drugs against certain reproductive cancers. Raloxifene (RAL) for example is an antiestrogen in the

reproductive tissues, but acts as a partial agonist in bone and lowers blood cholesterol. It is used clinically in the treatment of osteoporosis and hormone-dependent breast cancer.

5.1.2.3 THE LIGAND-BINDING POCKET

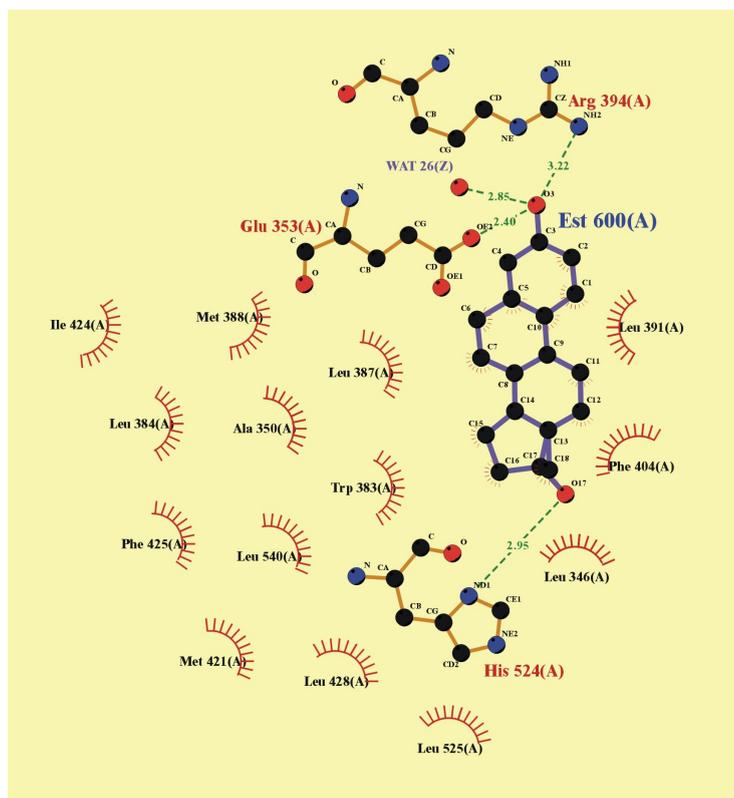


Fig. 5.4 Schematic representation of the interactions made by estradiol within the binding niche. Residues interacting with the ligand or lining the cavity are shown. Those making direct hydrogen bonds to the ligand are depicted in ball-and-stick style.

The estradiol binding cavity is located in the inner hydrophobic core of the protein and is formed by parts of H3 (Met342 to Leu354), H6 (Trp383 to Arg394), H8 and the preceding loop (Val418 to Leu428), H11 (Met517 to Met528), H12 (Leu537 to His547) and the S1/S2 hairpin (Leu402 to Leu410).

There are two polar ends in the pocket that make direct hydrogen bonds with the ligand. In the case of estradiol, the phenolic hydroxyl of the A-ring interacts with OE2 of Glu353, with the guanidinium group of Arg394 and a water molecule which is conserved for all crystal structures determined. The second polar end is His524, which interacts with the 17- β hydroxyl of the D-ring. The same H-bonding interactions are made in all other agonist structures of hER α , eg. in the complex with DES.

The rest of the binding niche is hydrophobic, consisting mainly of leucine residues, and participates in a number of hydrophobic contacts that are concentrated over the middle rather lipophilic part of the molecule.

5.1.3 WORKING PLAN

The aim of this project is the development of a new chemically inducible regulatory system on the basis of the hER α LBD. As the system is intended to be used for gene regulation in plants only, the designed ligands should not interact with other transcription systems (ie. those in mammals). To avoid this a specificity will be introduced by exchanging amino acids in the ligand binding pocket of the hER α LBD and constructing compounds that interact only with the mutants and not the native ligand-binding domain.

In this context the first step was to examine the docking of the ligands in the native LBD and upon that to modify both the protein binding niche and the molecules to a new specific system.

The organic ligands investigated in this thesis were obtained from Carsten Vock, from the group of Prof. Tietze, Institute of Organic Chemistry, University of Göttingen.

5.2 METHODS

5.2.1 MODELING

For the modeling studies the hER α -LBD complex with diethylstilbestrol (DES) and a GRIP1 NR Box II peptide was chosen from the existing crystal structures in the PDB (PDB ID: 3ERD). Reasons for this were that the structure has an agonistic conformation, it has an overall good quality (the resolution is high (2.0Å), the B-values and the R-factors are acceptable (mean B = 33.9 Å², R₁ = 0.196, R_{free} = 0.240), and the first copy (A) of the receptor in the asymmetric unit is uninterrupted in its main chain. Missing side chain atoms of the LBD copy A (AAs: 305-306, 330, 332, 335, 437, 460-467, 469-470, 531) were inserted using the internal rotamer library of insight2000.1. The terminal histidine 550 consisted of only one nitrogen atom, and was modeled as an N-terminus. Hydrogen atoms were geometrically calculated and added to all amino acids of the receptor. For all calculations the peptide and the DES molecule were removed.

In a first step the ligands were built up and geometrically optimized using the force field AMBER implemented in the program SYBYL. Subsequent docking of the ligands in the binding pocket of LBD-hER α was performed with the program FlexX. Using a radius of 4.5 Å of the active site proper solutions of the FlexX docking results were used for force field energy minimizations of the ligand-LBD complex with the CFF91 force field implemented in Insight2000.1/Discover2.98.

5.2.2 THE LIGANDS

The ligands used in the modeling studies were on the one hand designed to resemble in their structure and functionality the natural ligand of hER α , but on the other hand it is important that they can be varied in several ways, in order to fit to different mutations in the receptor. Therefore they were constructed so that they mimic the main properties of estradiol, but also bear additional organic moieties to interact with a mutated protein structure. Three different groups of ligands were designed by Carsten Vock, Institute of Organic Chemistry, University of Göttingen, shown in Table 5.2.

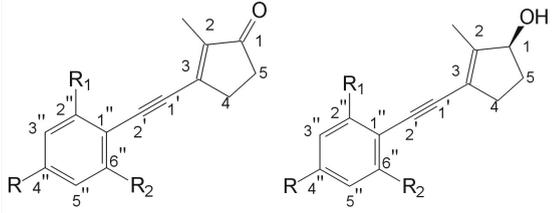
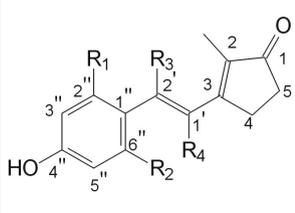
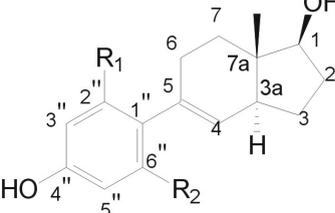
Ligands		
Group 1	Group 2	Group 3
		
R = OH, NH ₂ , CF ₃ , OCH ₃ R _{1,2} = Me, Et, iPr, Ph, OH	R _{1,2,3,4} = Me, Et, iPr, Ph, OH, CF ₃	R _{1,2} = Me, Et, iPr, Ph, OH, CF ₃

Table 5.2 Ligands designed by Carsten Vock in the group of Prof. Tietze.

In all cases the distance between the substituent R (in the case of groups 2 and 3 the hydroxyl) of the phenyl ring and the oxo group of the five-membered ring is similar to the distance between the 3-OH and the 17-OH group of the estradiol molecule (10.9 Å, Anstead et al., 1996). The estimated values for R = OH, R₁/R₂/R₃/R₄ = H is 11.4 Å for group 1 and 11.2 Å for group 2.

The first group of compounds consists of a phenyl moiety connected via a triple bond to a cyclopentenole or a cyclopentenone ring. Modification of the molecules can be achieved by modulating R, R₁ and R₂.

In the second group the triple bond is replaced by a double bond, allowing the introduction of two further substituents R₃ and R₄. The group *meta* to R₁ and R₂ is a hydroxyl group, as it was clear after modeling of the group 1 molecules that any other choice was not energetically favorable.

In the third group the phenyl moiety is connected via a single bond to a cyclohexene-cyclopentanole ring system. Only two variation possibilities are possible at the phenyl ring, namely R₁ and R₂ *meta* to the hydroxyl.

5.3 RESULTS AND DISCUSSION

5.3.1 MODELING STUDIES OF LBD WITH KNOWN AGONISTS

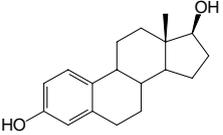
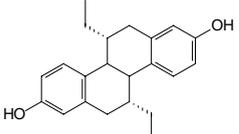
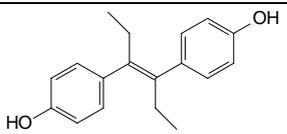
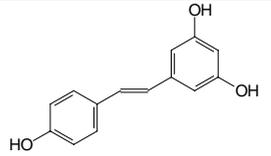
Ligand	Structure	Energy (kcal)	H-bonds in the binding pocket
Estradiol		-485013.8	Glu353, Arg394 His524
Tetrahydrochrysenene		-484991.5	Glu353, Arg394 His524
Diethylstilbestrol		-511176.1	Glu353, Arg394, His524
Resveratrol		-511187.1	Glu353, Arg394, His524

Table 5.3 Energies and interactions of calculated complexes with known agonists.

For comparative reasons the molecules estradiol, tetrahydrochrysenene (THC), diethylstilbestrol (DES) and resveratrol were docked in the pocket and the resulting complex

was energy minimized. In that way an energy range was defined for compounds that are proven to be agonists of the receptor.

Interestingly, the lowest energy is achieved by resveratrol. This phytoestrogen occurs naturally in grapes and red wine. It is a stilbene that in its *trans*-isomer form appears to have greater anticancer properties than in the *cis*. Resveratrol was recently proven to be an agonist for the estrogen receptor (Gehm et al, 1997), but no structural information on its complex with the hER α LBD is known.

The starting coordinates for resveratrol used were obtained from its crystal structure shown in Fig. 5.5 (Kubas, 2003). The molecule is almost planar in its structure and the distances between the oxygens of the two phenyl moieties are 11.31 Å (O1-O2) and 10.75 Å (O1-O3).

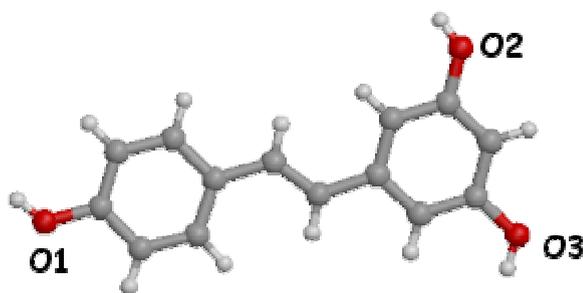


Fig. 5.5 Crystal Structure of Resveratrol, determined by M. Kubas.

Resveratrol was fitted into the hER α LBD binding pocket with FlexX, and the energy of the complex was minimized. The resveratrol-hER α complex (-511187.1 kcal) is even slightly more favorable than the original ligand of the X-ray structure DES (-511176.1 kcal). This could be a result of the high structural similarity of the two compounds. The binding niche of the final model is shown in Fig. 5.6.

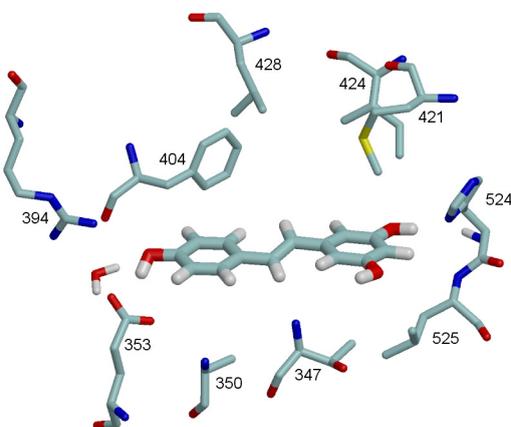


Fig. 5.6 Binding pocket of the energy minimized structure of hER α LBD with resveratrol.

The molecule lies planar inside the binding pocket. The A-phenyl ring of resveratrol is oriented towards Arg394 and Glu353 with O1 corresponding to the 3-OH of estradiol. Hydrogen bonds are formed to the oxygen atom of Glu353 (O...O: 2.84 Å) and Arg394 (N...O: 3.32 Å). The B-phenyl ring of resveratrol interacts with His524 via one of its hydroxyl groups (N...O: 2.83 Å), while the second hydroxyl group is pointing towards Thr347, but is not involved in any hydrogen bonding.

5.3.2 MODELING STUDIES OF THE LIGANDS WITH THE LBD

5.3.2.1 GROUP 1

All minimized complexes are listed in Table 5.4.

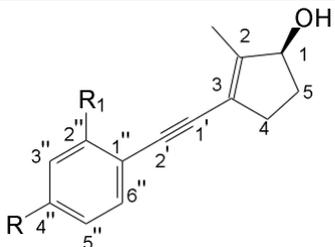
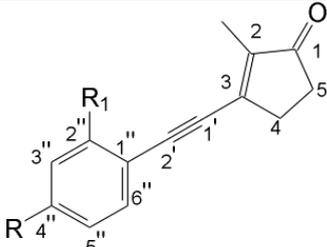
									
	R	R ₁	Energy (kcal)	H-bonds in the pocket		R	R ₁	Energy (kcal)	H-bonds in the pocket
1	OH	/	-511194.6	Glu353, Arg394, His524	5	OH	/	-511183.6	Glu353, Arg394
2	OH	Me	-485025.4	Glu353, Arg394, His524	6	NH ₂	/	-484949.4	Glu353, Arg394, Leu387
3	OH	Et	-485053.0	Glu353, Arg394, His524	7	OH	Me	-485003.7	Glu353, Arg394
4	OH	/	-485012.1	Glu353, Arg394, His524	8	OH	Et	-484999.8	Glu353, Arg394

Table 5.4 Energies and interactions of the calculated complexes for group 1.

In all cases (except **4**) the structures were docked in a way that the phenyl moiety corresponds to the A-ring of estradiol in hER α . In this area of the binding niche not much space is left and only small substituents in the position R₁ are allowed. The hydroxyl group at the cyclopentenole ring allows the docking of the ligand in a reverse mode, with the pentenole-OH hydrogen bonded to Glu353 and Arg394 and the phenyl-OH group interacting with His524. This was the case for structure **4**. Nevertheless, the energetic difference of 26182,5 kcal indicates a rather unlikely formation of this complex.

Due to low steric hindrances with the amino acids of the active site the unsubstituted structures **1** and **5** are energetically by far the most favorable. In the list of the substituted compounds the biggest substitution at R₁ is an ethyl group. Already the isopropyl moiety

makes the ligand too bulky to be docked in the binding pocket by FlexX. Nevertheless, for the pentenoles, the energy of the ethyl substituted ligand (**3**) is lower than that of the methylated compound (**2**), probably owing to a nice cavity in the area of amino acids Met388, Leu428, Met421 and Ile424. In both cases the hydrophobic groups Me and Et are accommodated towards this direction (Fig. 5.7).

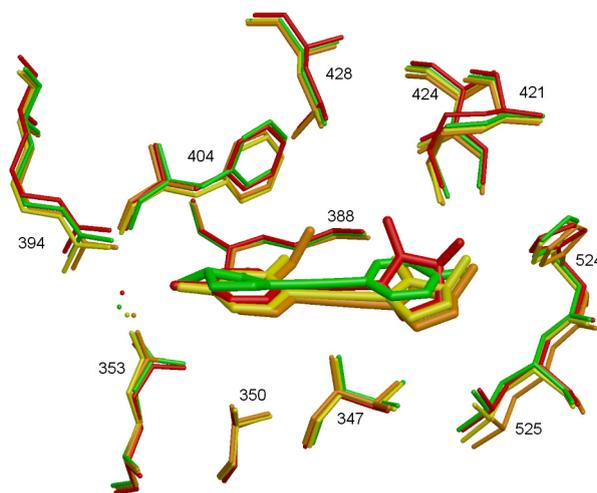


Fig. 5.7 Superposition of the binding pockets of the minimized structures of hER α with **1** (red), **2** (yellow), **3** (orange) and **4** (green).

The hydrogen bond of the OH-group in the cyclopentenoles with His524 is no longer possible when changing to cyclopentenones. Furthermore, the conjugated system in the five membered ring in the pentenones forces all carbon atoms to lie in one plane with the keto group, making the whole ring sterically smaller than in the pentenoles. An additional consequence is that the unbound pentenone ring is more flexible and can be rotated more easily around the single bond C3-C1' in a way that sterical hindrances with R₁ and R₂ are minimized. For compounds **7** and **8** R₁ is turned towards Met388. For compound **7** the cyclopentenone ring is rotated 49.5° to the phenyl moiety to avoid interactions between the methyl R₁ and the methyl of the pentenone. Interestingly, in the case of **8** the ring systems are almost in the same plane (-6.5°) because otherwise the bulky isopropyl group would interact with the ethyl group (Fig. 5.8).

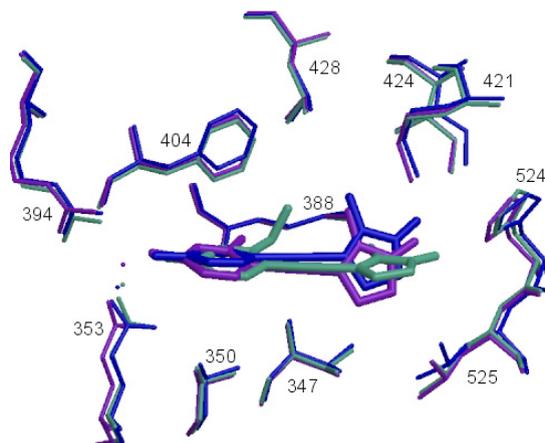


Fig 5.8 Superposition of the binding pockets of the minimized structures of hER α with **5** (blue), **7** (purple) and **8** (green).

In **6** the phenyl OH group is substituted with an amine group. The H-bonding acceptor for the second hydrogen atom of the nitrogen is the main chain carboxyl of Leu387, with a somewhat long H...O distance of 2.8 Å. In spite of this additional hydrogen interaction the energy is quite high.

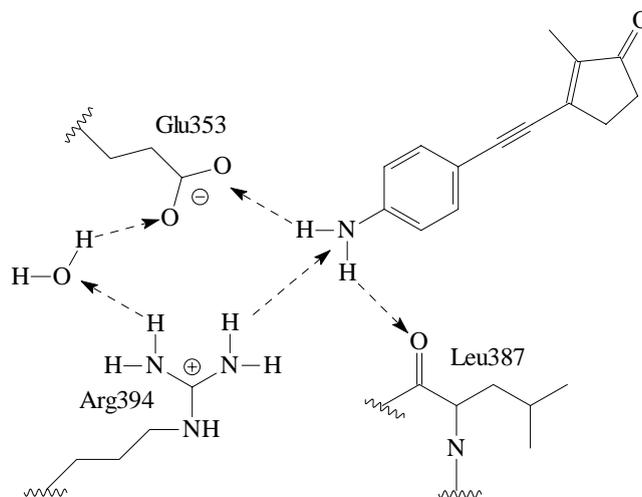


Fig. 5.9 Schematic drawing of the interactions between compound **6** and amino acids of the binding pocket of hER α .

5.3.2.2 GROUP 2

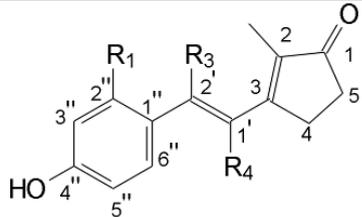
					
	R ₁	R ₃	R ₄	Energy (kcal)	H-bonds in the pocket
9	/	/	/	-511160.2	Glu353, Arg394, His524
10	/	Me	/	-511097.2	Glu353, Arg394, His524
11	/	Et	/	-484923.4	Glu353, Arg394, His524
12	/	iPr	/	-484950.2	Glu353, Arg394, His524
13	/	CF ₃	/	-485038.5	Glu353, Arg394, His524
14	Me	Me	Me	-484949.9	Glu353, Arg394, His524

Table 5.5 Energies and interactions of the calculated complexes for group 2.

In the second group of compounds (Table 5.5) the conformational flexibility is in general higher than for group 1, owing to the additional substituents at the double bond. Nevertheless, the molecules are longer and “bigger”, and the overall energy values slightly higher. The interaction pattern is very similar to the first group, with the phenyl OH hydrogen-bonded to NH₂ of Arg394 and OE1 or OE2 of Glu353, and the cyclopentenole hydroxyl interacting with His524.

The unsubstituted compound (**9**) is again (as for group 1) the most favorable ligand, but the energetic difference to the compound methylated in position R₃ (**10**) is much smaller than in group 1. Interestingly, ligands with none or small substituents at position R₃ (compounds **9**, **10** and **14**) show a different conformation than those carrying bulkier groups (compounds **11**, **12** and **13**). In the first case this group points towards Met388. By introducing an ethyl group at this position, the molecule twists around turning the bulky moiety towards amino acids Leu347, Ala350, Trp383 and Leu384, which are forced to move away in order to avoid steric clashes with the ligand (e.g. in **12**, shown in Fig. 5.10 in grey).

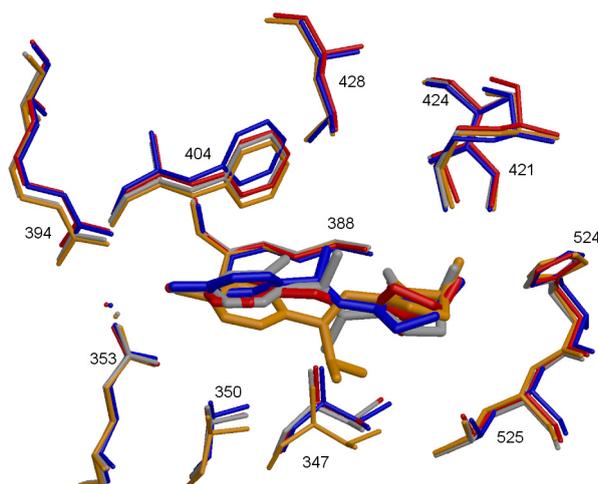


Fig. 5.10 Superposition of the binding pockets of the minimized structures of hER α with **9** (red), **10** (blue), **12** (grey) and **14** (orange).

The same is true for the isopropyl and trifluoromethyl substituents. In general, with bulkier substituents the overall energy is getting higher, thus these compounds are less favorable.

5.3.2.3 GROUP 3

	R ₁	Energy (kcal)	H-bonds in the pocket
15	/	-484991.4	Glu353, Arg394, His524
16	Me	-485015.9	Glu353, Arg394, His524
17	Et	-485046.1	Glu353, Arg394, His524
18	CF ₃	-484970.0	Glu353, Arg394, His524
19	iPr	-484934.0	Glu353, Arg394, His524

Table 5.6 Energies and interactions of the calculated complexes for group 3.

Ligands belonging to the third group have the closest structural similarity to steroids. With the assignment of the phenyl moiety to ring A of estradiol, one can easily recognize the second part of the structure as the slightly modified rings C and D of the hormone. Due to the absence of ring B, the molecule is less rigid than estradiol. Whereas estradiol is in a

plane inside the binding pocket, in all compounds of the third group the “C, D” moiety is almost perpendicular to the phenyl ring in the minimized complexes. The angles between the two moieties vary between 80.4° and 89.1°. This flexibility in the ligands allows the accommodation of bulky substituents in the *meta* position to the phenyl hydroxyl with minimal steric interactions. Only small energetic differences are observed between the unsubstituted compound **15** and the substituted structures. The biggest difference of 112 kcal is found between **19** (isopropyl substitution) and **17** (with an ethyl group at R₁), where **17** is the most favorable (Table 5.6).

The superposition of the structures reveals a similar conformation for all compounds (Fig. 5.11).

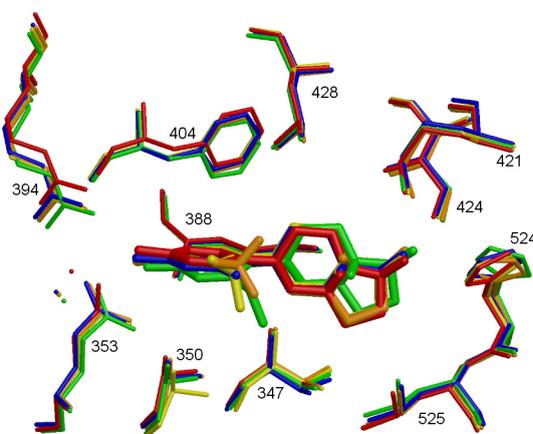


Fig. 5.11 Superposition of the binding pockets of the minimized structures of hER α with **15** (red), **16** (blue), **17** (green) and **19** (yellow).

While the C,D moiety is perpendicular to the phenyl ring, R₁ and the C7a methyl group have to move away from M388, towards amino acids Thr347, Leu346 and Met343.

5.3.3 MODELING STUDIES OF MUTATED LBDS WITH THE LIGANDS

5.3.3.1 POSSIBLE MUTATION SITES

Three major directions were considered for mutations of the ligand binding pocket.

Firstly, bulky amino acids are introduced into the binding niche, so that the overall volume is decreased and only small compounds can bind to the receptor. The aim is to achieve a better fit and thereby a higher affinity of the ligands to the mutated protein than to the native one. However, since there are already voluminous residues in the pocket (Leu, Ile, Phe, Trp,

Arg, Glu), it is difficult to make them larger by mutation. A further undesired consequence would be that this mutated hER α has also a high affinity to estradiol. To avoid side effects, the compound should only bind to the mutated structure.

Secondly, to increase the space in the binding pocket, bulky amino acids should be replaced by smaller ones. As a consequence big compounds that do not fit in the native receptor would be accommodated in the mutant. This possibility was actually achieved in the molecular modeling studies (mutations M388A/L428A, M388A/L428A/M421A, W383A/L384A). Results will be shown later on in this thesis.

Thirdly, a change in the hydrophobicity of the pocket is introduced by replacing lipophilic residues with hydrophobic ones. The ligands are then modified by substitution with polar groups (i.e. hydroxyl) that can interact with the mutated amino acids. One should be rather precautionous with this mutation, as the nature of the whole pocket is modified and could lead to a different conformation. One mutation was investigated in this direction (M388T).

The amino acid residues to be mutated are shown in Fig. 5.12.

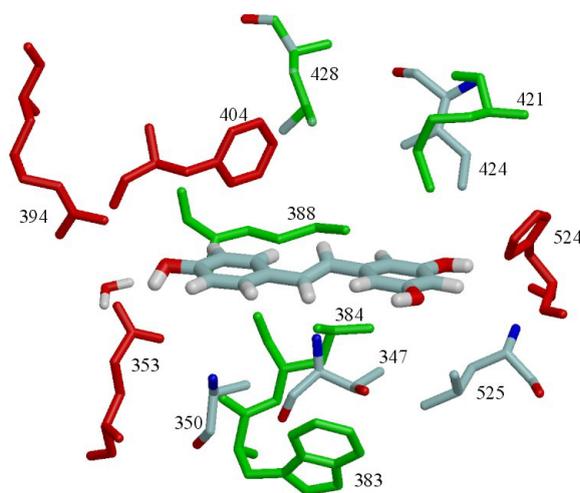


Fig. 5.12 Possible mutation sites. The ligand binding pocket of hER α LBD is shown in complex with resveratrol. Residues that should not be changed are shown in red, residues that were mutated in the modeling studies are shown in green.

The residues that should remain unchanged are the ones of the two polar ends of the cavity, which serve as anchoring points for the molecules: Arg394 and Glu353 that perform hydrogen bonds with the hydroxyl group of the phenyl moiety and His524 that interacts with the hydroxyl group of the cyclopentenole/cyclopentanole ring of the ligands. Furthermore, the phenylalanine 404 was not replaced as it is involved in π -stacking with the phenyl ring of the compounds.

The docking and energy minimization studies of the native complexes showed different results for the three groups of ligands:

a) All substituents at R₁ of group 1 point towards the direction of Met388. By replacing this long chain with a smaller amino acid it might be possible to introduce isopropyl or even phenyl moieties at this position of the ligands. Nevertheless, this single mutation was not sufficient to fit bulkier compounds into the pocket. In the same area are the residues Leu428 and Met421. Their subsequent mutation to alanines allowed the accommodation of big molecules with up to two substituents on the phenyl moiety.

b) As regards group 2, the two additional substituents on the double bond seem to prefer the area of Leu346, Leu384 and Trp383. However, the reduction of Leu346 to an alanine had no effect on the docking. Successful was the mutation L384A/W383A, that facilitated the fitting of symmetrically substituted ligands of the second group in the pocket.

c) For the third group it was not possible to find a suitable mutation.

Interestingly, none of the mutations applied in this work were found in NucleaRDB, a Nuclear Receptor Data base which collects mutational data from the literature.

5.3.3.2 THE MUTATION M388A/L428A

As mentioned above only compounds of group 1 were affected by this mutation. Aiming at the accommodation of bulkier molecules in the binding niche, the two amino acids M388 and L428 were mutated to alanines. This double mutation allows a wider variation of R₁ towards bigger substituents like phenyl groups, and additional substitution at position 6'' is also possible. Ligands that were docked and minimized successfully in the mutated binding pocket are shown in Table 5.7 (compounds **20** to **23**).

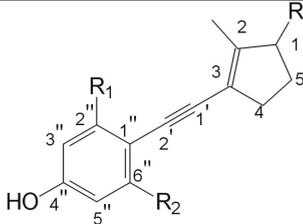
					
	R	R ₁	R ₂	Energy (kcal)	H-bonds in the pocket
20	Keto	Ph	/	-484809.8	Glu353, Arg394
21	Keto	Ph	Me	-484887.7	Glu353, Arg394
22	Keto	Ph	Et	-484805.1	Glu353, Arg394
23	Hydroxyl	Me	Et	-484967.2	Glu353, Arg394, His524

Table 5.7 Energies and interactions of the calculated complexes for group 1 and for the mutation M388A/L428A.

Owing to the rotational flexibility of the cyclopentenone ring steric clashes of its methyl group with bulky substituents of the phenyl moiety can be avoided for compounds **20** to **22**. While R₁ is accommodated in the cavity created by the substitution by the rather long Met388 to a much smaller Ala and of the Leu428 to Ala, the methyl group of the five-membered ring is turned in the opposite direction (Fig. 5.13). However, in that way R₂ is restricted to groups not bigger than ethyl substituents.

In the case of compound **23** the cyclopentenole is fixed in its orientation due to the hydrogen bonding of its hydroxyl to His524. Only small rotations of the five-membered ring around the C3-C1' bond are observed that do not cause steric clashes with Met421 (angles between the two rings **20**: 26.9°, **21**: 19.6°, **22**: 16.1° and **23**: 17.9°). Thus, the methyl group of the cyclopentenole points towards the new cavity, making a substitution of groups bigger than ethyl impossible.

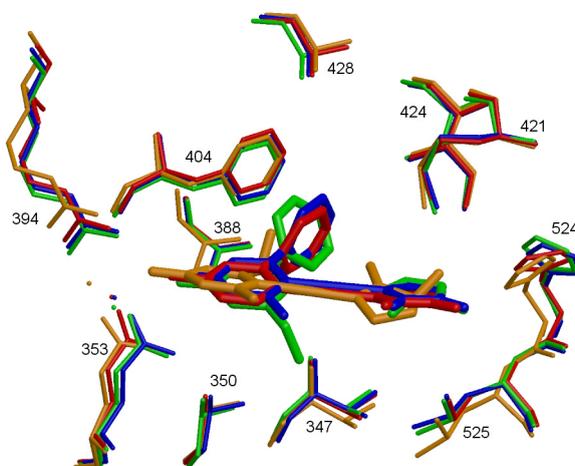
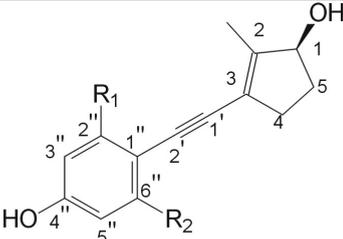


Fig. 5.13 Superposition of the binding pockets of the minimized structures of hER α mutant M388A/L428A with **20** (red), **21** (blue), **22** (green) and **23** (orange).

5.3.3.3 THE MUTATIONS M388A/L428A/M421A AND M388A/L428A/M421I



	R ₁	R ₂	Energy (kcal) M388A/L428A/M421A	Energy (kcal) M388A/L428A/M421I	Interactions in the pocket
24	Me	iPr	-484942.9	-485011.2	Glu353, Arg394, (His524)
25	Me	Ph	-484877.0	-484857.2	Glu353, Arg394, (His524)
26	Et	Et	-484986.8	-484934.2	Glu353, Arg394, (His524)
27	Et	iPr	-484995.8	/	Glu353, Arg394, His524
28	Et	Ph	-484891.1	-484917.5	Glu353, Arg394, His524
29	iPr	Ph	-484899.5	/	Glu353, Arg394, His524
30	/	Ph	-484858.0	-484900.0	Glu353, Arg394, (His524)

Table 5.8 Energies and interactions of the calculated complexes for group 1 and for the mutations M388A/L428A/M421A and M388A/L428A/M421I.

Again, only molecules of group 1 are affected by this triple mutation in the binding pocket (Table 5.8, compounds **24** to **30**). The additional substitution of Met421 to Ala makes it possible for the cyclopentenole ring to rotate to an almost perpendicular position to the phenyl moiety, without any clashes of its methyl group with side chains in the region of amino acids Ala428, Ala421 and Ile424.

The energy calculated for the complexes is somewhat higher compared with the monosubstituted ligands docked in the wild type ER α . A slight trend that the phenylsubstituted molecules are less favorable (energy difference of 100 kcal) can be observed.

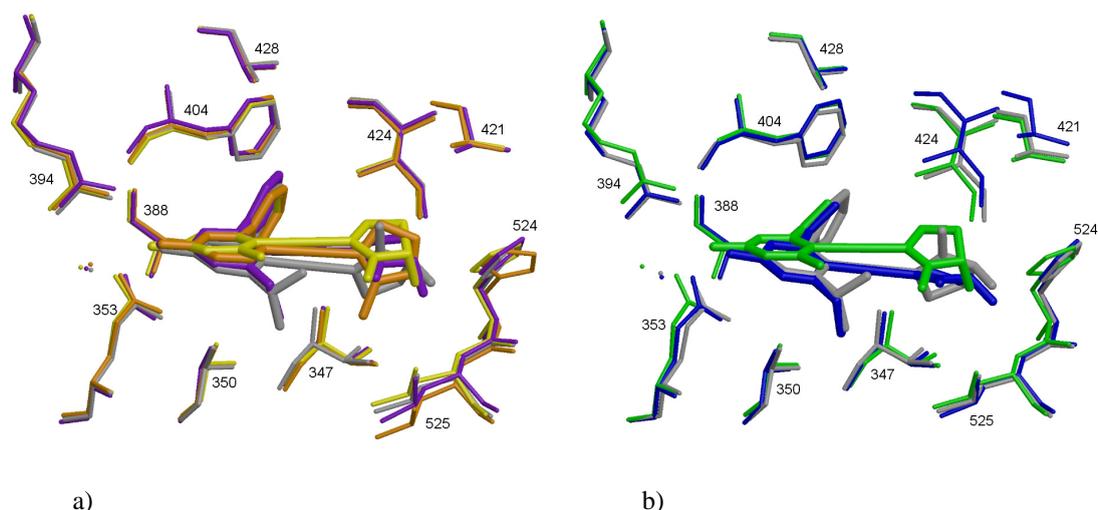


Fig. 5.14 a) Superposition of the binding pockets of the minimized structures of hER α mutant M388A/L428A/M421A with the phenylsubstituted ligands **30** (orange), **25** (yellow), **28** (purple) and **29** (grey) b) Superposition of the binding pockets of the minimized structures of hER α mutant M388A/L428A/M421A with the isopropylsubstituted ligands **24** (green), **27** (blue), and **29** (grey).

In all calculations the bulkiest substituent is turned towards the mutated amino acids Ala388 and Ala428. All phenyl substituted compounds (**25**, **28**, **29** and **30**) are shown in Fig. 5.14a. The bigger the second substituent ($H > Me > Et > iPr$) becomes, the more the cyclopentenole avoids intramolecular steric clashes of its methyl group by rotating around the triple bond (**30**: -70.5° , **25**: -18.5° , **28**: -8.7° , **29**: 76.9°). The biggest movement is observed in the case of the bulky isopropyl in R₂ (compound **29**). The protein structure reacts as well by moving His524, and thereby following the rotation of the cyclopentenole. The same trend is observed in the superposition of the isopropyl substituted ligands **24**, **28** and **29** (Fig. 5.14b). As long as the second substitution is small (methyl or ethyl), the *i*-propyl group is accommodated towards Ala388. As soon as the bulky phenyl group is introduced, the positions of the two groups are exchanged. These variations can be followed in the movement of the cyclopentenole.

An alternative triple mutation studied was M388A/L428A/M**421I**. Aim was the investigation of the effects of a mutation to an amino acid of medium size and not as small as an alanine. The exchange of Met421 to Ile is a reasonable one as it already exists in nature (difference ER α \rightarrow ER β), causing no considerable change of conformation in the binding pocket.

The mutation M421I allowed the docking of compounds **24**, **25**, **26**, **28** and **30** in the binding pocket. The bulky ligands with both phenyl and ethyl or isopropyl groups (**27** and **29**) could

not be docked in the binding niche. The energy values are in the same range as for the M421A mutation.

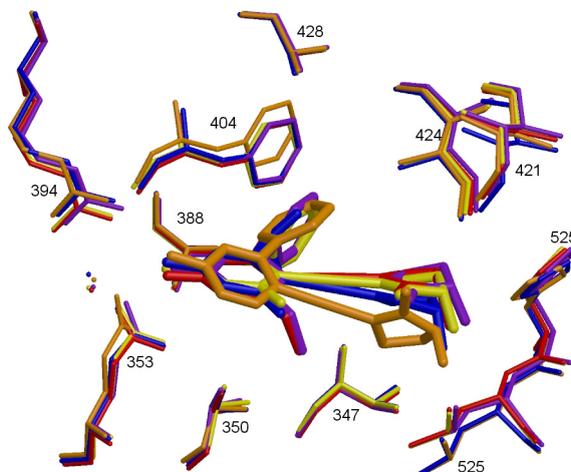


Fig. 5.15 Superposition of the binding pockets of the minimized structures of hER α mutant M388A/L428A/M421I with the ligands **24** (blue), **25** (yellow), **26** (red), **28** (purple) and **30** (orange).

The cavity in the area of Ile421 is not as big as in the case of M421A, therefore the methyl group of the cyclopentenole is turned towards Leu525. This residue follows the movement of the ring. The nearer the five-membered ring, the more Leu525 has to turn away (Fig. 5.15). Freedom in movement is also observed for the phenyl moiety of the ligands. The hydroxyl group as well as the bulky substitution at R₂ force Arg394 and Phe404 to change their position dependent on the ligand. Interestingly, the dispositions are the strongest for the case of compound **30**, which is monosubstituted with a phenyl group.

Another surprising feature is that only the bulkiest compound **28** performs hydrogen bonding with its cyclopentenole hydroxyl to His524. In all other cases, the OH group is not involved in any interactions, but is oriented towards Leu525.

5.3.3.4 THE MUTATION M388T

The mutation of Met388 into a threonine comes along with a change in the hydrophobicity of the binding niche. The addition of a further H-bond partner in the pocket allows hydrophilic moieties like hydroxyl for all groups of ligands (Table 5.9, compounds **31** to **34**).

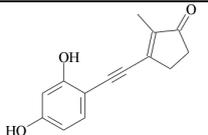
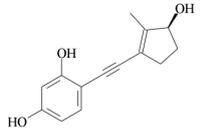
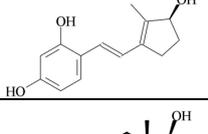
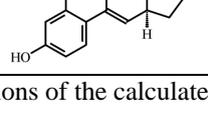
	Structure	Energy (kcal)	H-bonds in the pocket
31		-484988.3	Glu353, Arg394, Thr388
32		-484991.4	Glu353, Arg394, Thr388, His524
33		-484994.1	Glu353, Arg394, Thr388, Gly521
34		-484998.7	Glu353, Arg394, Thr388, Gly521

Table 5.9 Energies and interactions of the calculated complexes for the mutation M388T.

In the docked protein-ligand complexes an additional hydrogen bond from O γ 1 of the threonine acting as a hydrogen acceptor to the hydroxyl attached at C2'' of the compounds is observed. The distance of the atoms involved (O γ ...H) vary from 2.1 Å in compound **31**, over 2.6 Å for compounds **33** and **34** to the very long distance of 3 Å in the case of compound **32**. The reason for this long H-bond is the fact that the molecule possesses three OH groups, which are all involved in hydrogen bonding. The orientation providing the best fit is the one where all of them form weak hydrogen bonds with long distances. Thr388 is further stabilized by an intramolecular interaction to the main chain carbonyl of Leu384.

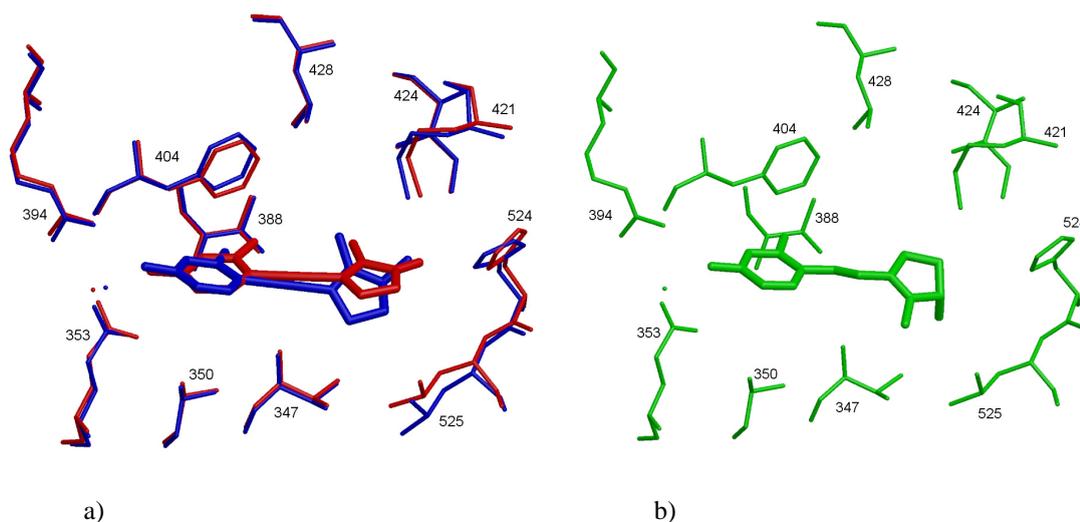


Fig. 5.16 a) Superposition of the binding pockets of the minimized structures of hER α mutant M388T with the ligands **31** (blue) and **32** (red) b) Binding pocket of the minimized structure of hER α mutant M388T with ligand **33** (green).

In the case of compound **33** the hydroxyl group of the cyclopentenole is hydrogen bonded to the main chain carbonyl O-atom of Gly521 instead of His524 with a distance of 3.1Å (O...Oyl). The methyl group is oriented towards Thr347.

5.3.3.5 THE MUTATION W383A/L384A

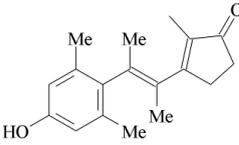
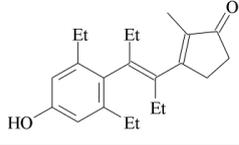
	Structure	Energy (kcal)	H-bonds in the pocket
35		-485038.8	Glu353, Arg394
36		-485079.2	His524

Table 5.10 Energies and interactions of the calculated complexes for the mutation W383A/L384A.

This mutation affected the ligands of group 2 and facilitated the docking of symmetrically substituted compounds in the binding niche (Table 5.10). While compound **35** interacts via its phenyl hydroxyl group with amino acids Arg394 and Glu353, compound **36** is docked the other way around in the pocket (Fig. 5.17).

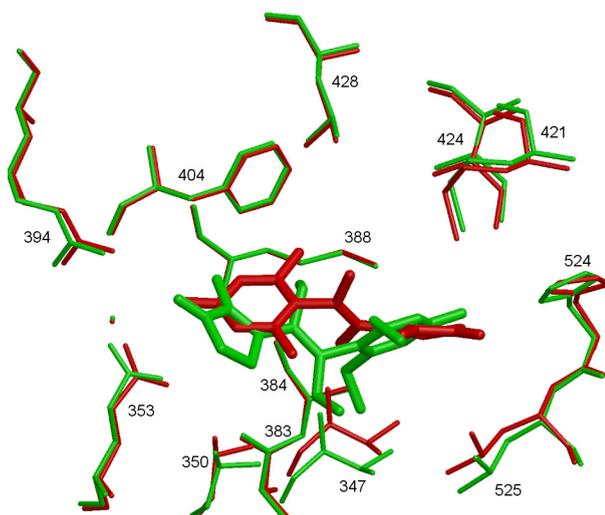


Fig. 5.17 Superposition of the binding pockets of the minimized structures of hER α mutant M388A/L428A/M421I with the ligands **35** (red) and **36** (green).

The phenyl-OH of **36** is hydrogen bonded to His524 with a H...N γ distance of 2.4 Å while the keto group of the cyclopentenone is interacting weakly with Arg394 (O...H distance 2.9 Å).

The ethyl substituents of the phenyl moiety are oriented towards residues Leu525 and Leu347, which move away to avoid steric clashes with the bulky groups, while the ethyl groups of the double bond are positioned towards Met388 and in the cavity created by the mutation of residues Trp383 and Leu384.

SUMMARY AND CONCLUSIONS

6.1 THE 7-AAMD-SSDNA COMPLEX

The binding of 7-AAMD to the seemingly single-stranded DNA oligomer 5'-TTAG[Br⁵U]T-3' has been studied via X-ray crystallography. The crystallization of non-complementary DNA with AMD has not been reported before. Instead of the expected 1:1 ratio of drug to ssDNA an antiparallel mismatched double stranded DNA complexed by the 7-AAMD was found in the crystals.

The intercalation of 7-aminoactinomycin D between mismatched G-BrU or G-T steps constitutes an unprecedented structure. One possible explanation is the recognition of such mismatches by actinomycin. This would require the existence of the non-complementary DNA in a double stranded form before the drug binds the DNA. Nevertheless, thermally resolved absorption spectra of non-complementary oligonucleotides in the presence and absence of 1:1 concentrations of drug:strand provide no evidence for secondary structures over a wide temperature range, but indicate that the DNA remain single-stranded when AMD is bound.

A further possible explanation is the strong preference of actinomycin for G-G steps, independent of their binding partners, which forces the mismatching of the strands. In all X-ray or NMR structures of AMD complexed with DNA (double-stranded or "single-stranded") that are deposited in the Protein Data Bank, the phenoxazone ring is sandwiched between GG steps, with a dominating overlap of the guanine rings over the chromophore. NMR studies of a 1:1 5'-TTAGTT-3':AMD solution carried out by the group of Prof. Jovin revealed strong NOE correlations between the protons H1' of the guanine and the methyl protons of the D-threonine of actinomycin D and weak correlations between the H2'1, H2'2 of the thymine and H_A of the AMD D-proline. These findings are in good agreement with the interactions in the crystal structure (Fig. 6.1).

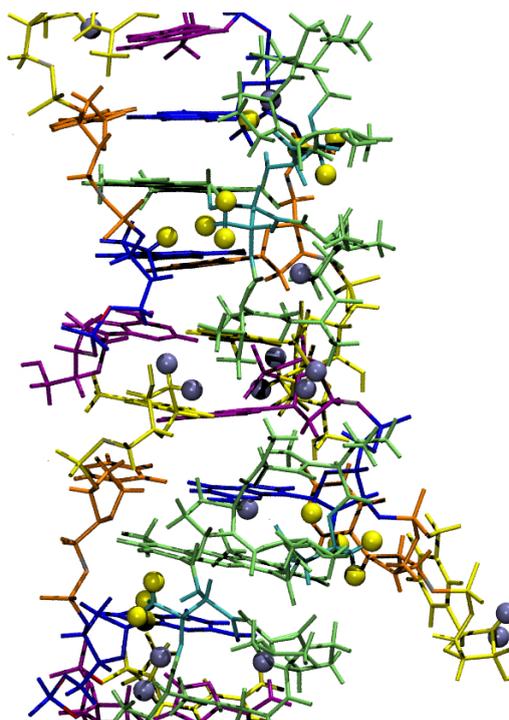


Fig. 6.1 Part of the hexagonal structure with highlighted protons showing NMR correlations. Protons showing a strong NOE correlation are shown in yellow (H1'_GUA, Me_DTH), those showing a weak correlation are shown in blue (H2'1, H2'2_THY, H_A_DPR). Guanines are shown in blue, thymines in yellow, adenines in red, bromouracils in orange and the actinomycins in green.

Molecular dynamics simulations were performed upon these studies by Reinhard Klement, MPI for Biophysical Chemistry, Göttingen, using a 1:1 complex of drug (AMD) to strand (5'-TTTAGTTT-3') with the phenoxazone intercalated in the G-T step as a model (Fig. 6.2). In the resulting structure the phenoxazone ring is positioned parallel to the guanine ring, so that a maximum overlap is achieved. This π -stacking allows very little movement for the two rings, while the flanking thymines and the adenine are free to rotate in the solvent cage.

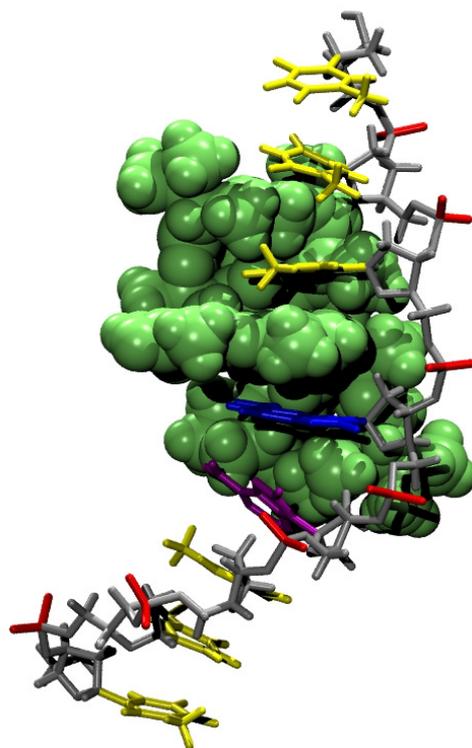


Fig. 6.2 Snapshot of the MD simulation of the 5'-TTTAGTTT-3' complexed with AMD. Guanine is shown in blue, adenine in purple and thymines in yellow.

The only deposited structure of a non-complementary DNA complex is the NMR model of 5'-CCGTT₃GTGG-3' with actinomycin D (PDB ID: 1OBF). This oligomer forms a TTT-hairpin structure with tandem G-T mismatches, with the phenoxazone ring sandwiched between them. In this case the six-membered rings of the guanines and the phenoxazone were found to stack almost on top of each other, while the bases of their respective binding partners (thymines) are completely destacked from the hydrophobic core.

Coordinates and observed structure factor amplitudes have been deposited in the Protein Data Bank (accession codes: hexagonal structure 1UNJ, orthorhombic structure 1UNM).

6.2 THE D,L-ALTERNATING PEPTIDE

All of the known alternating non-cyclic d,l-peptide structures show the double stranded antiparallel β -helix motif. The crystallization of the peptides was performed in organic solvents. A comparison of the present structure with them is therefore difficult. Many of the structural details of the H-(L-Tyr-D-Tyr)₄-Lys-OH, like crystal packing, can be explained by the connecting water network, which is not existent in the other structures. Nevertheless, in all of the cases the amino acid side chains are on the same side of the β -sheets, oriented towards the solvent. In that way, an inner channel is formed, in most of the cases big enough to allow the transfer of ions. Theoretically, it could be possible for the H-(L-Tyr-D-Tyr)₄-Lys-OH to form ion conductors.

Coordinates and observed structure factor amplitudes have been deposited in the Protein Data Bank (accession code 1UNO).

6.3 MODELING STUDIES ON hER α LBD

The investigation of the binding of organic ligands within the LBD of hER α led to five suggestions for mutations of residues in the binding pocket. In general it seems that compounds of the first group fit better in the binding niche of the protein, as they showed the best results. Ligands of this group carrying various substituents were docked in all mutations, resulting in chemically and energetically reasonable solutions.

In order to validate the theoretical studies performed in this thesis, the ligands discussed should be synthesized and characterized as regards their stability, their water solubility and their toxicity. Furthermore, the mutants suggested should be created. It is important that their conformation is that of the native agonistic hER α LBD. Then, the ligands should be tested in their affinity towards both the native and the mutated hER α LBD.

A further important aspect that should be investigated is the detailed energetical examination of the binding pocket. For the structures presented here the energy of the whole protein-ligand complex in a water cage was minimized. It would be interesting to focus on the energetic contribution of the occupied binding pocket in order to gain more precise information as regards the binding affinity.

REFERENCES

- Alexopoulos, E. (2001): Röntgenstrukturuntersuchungen an Vinamidinmetallkomplexen, sowie Kristallisation eines 7-Aminoactinomycin-ssDNA Komplexes, *Diplomarbeit*.
- Anstead, G.M., Carlson, K.E., Katzenellenbogen, J.A. (1996): The estradiol pharmacophore: Ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site, *Steroids*, **62**, 268-303.
- Aoyama, T., Dong, C.-H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., Morelli, G., Chua, N.-H. (1995): Ectopic expression of the *Arabidopsis* transcriptional activator Athb-1 alters leaf cell fate in tobacco, *Plant Cell*, **7**, 1773-1785.
- Böhm, H.-J. (1992a): The computer program LUDI: A new method for the de novo design of enzyme inhibitors, *J. Comput.-Aided Mol. Design*, **6**, 61-78.
- Böhm, H.-J. (1994): The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure, *J. Comput.-Aided Mol. Design*, **8**, 243-256.
- Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., Moras, D. (1995): Crystal structure of the ligand-binding domain of the human nuclear receptor RXR α , *Nature*, **378**, 681-689.
- Brünger, A.T. (1992): Free R value: A novel statistical quantity for assessing the accuracy of crystal structures, *Nature*, **355**, 472-475.
- Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.E., Bonn, T., Engström, O., Öhman, L., Greene, G.L., Gustafsson, J.-A., Carlquist, M. (1997): Molecular bases of agonism and antagonism in the oestrogen receptor, *Nature*, **389**, 753-758.
- Chan, W.C., White P.D., New York (2000), Oxford University Press Inc.
- Chin, K.-H., Chen, F.-M., Chou, S.-H. (2003): Solution structure of the ActD-5'-CCGTT₃GTGG-3' complex: drug interaction with tandem GT mismatches and hairpin loop backbone, *Nucleic Acids Res.*, **31** (10), 2622-2629.
- Chou, S.-H., Chin, K.-H., Chen, F.-M. (2002): Looped out and perpendicular: Deformation of Watson-Crick base pair associated with actinomycin D binding, *PNAS*, **99** (10), 6625-6630.
- Cowtan, K. (1994): 'dm': An automated procedure for phase improvement by density modification, *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34-38.
- Dauter, Z., Dauter, M., Dodson, E. (2002): Jolly SAD, *Acta Cryst. D*, **D58**, 494-506.
- Di Blasio, B., Benedetti, E., Pavone, V., Pedone, C., Spiniello, O., Lorenzi, G.P. (1989): Regularly alternating L,D-peptides. I. The double-stranded left-handed antiparallel β -helix in the structure of Boc-(L-Val-D-Val)₄-OMe, *Biopolymers*, **28**, 193-201.

- Di Blasio, B., Benedetti, E., Pavone, V., Pedone, C., Gerber, C., Lorenzi, G.P. (1989): Regularly alternating L,D-peptides. II. The double-stranded left-handed antiparallel β -helix in the structure of Boc-(L-Phe-D-Phe)₄-OMe, *Biopolymers*, **28**, 203-241.
- Dodson, E. (2003): Is is jolly SAD?, *Acta Cryst. D*, **D59**, 1958-1965.
- Eiler, S., Gangloff, M., Duclaud, S., Moras, D., Ruff, M. (2001): Overexpression, Purification and Crystal structure of Native ER α LBD, *Protein Expression and Purification*, **22**, 165-173.
- Ennifar, E., Carpentier, P., Ferrer, J.-L., Walter, P., Dumas, P. (2002): X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing, *Acta Cryst. D*, **D58**, 1262-1268.
- Evans, G., & Pettifer, R.F. (2001): *CHOOCH*: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra, *J. Appl. Cryst.*, **34**, 82-86.
- Gangloff, M., Ruff, M., Eiler, S., Duclaud, S., Wurtz, J.M., Moras, D. (2001): Crystal structure of Mutant hER α Ligand-binding domain reveals key structure features for the mechanism of partial agonism, *Journal of Biological Chemistry*, **276** (18), 15059-15065.
- Gehm, B.D., McAndrews, J.M., Chien, P.-Y., Jameson, J.L. (1997): Resveratrol, a polyphenolic compound found in grapes and wine, is an agonist for the estrogen receptor, *Proc. Nat. Acad. Sci. USA*, **94**, 14138-14143.
- Greene, G.L., Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., (1998): The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen, *Cell*, **95**, 927-937.
- Guo, J., Wu, T., Bess, J., Henderson, L.E., Levin, J.G. (1998): Actinomycin D inhibits human immunodeficiency virus type 1 minus-strand transfer in In Vitro and endogenous reverse transcriptase assays, *Journal of Virology*, **72** (8), 6716-6724.
- Hendrickson, W.A., Smith, J.L. & Sheriff, S. (1985): Direct phase determination based on anomalous scattering, *Methods in Enzymology*, **115**, 41-54.
- Hou, M.-H., Robinson, H., Gao, Y.-G., Wang, A.H.-J. (2002): Crystal structure of actinomycin D bound to the CTG triplet repeat sequences linked to neurological diseases, *Nucleic Acids Res.*, **30** (22), 4910-4917.
- Insight II/Discover 2.98 (MSI Inc.; San Diego, CA), 2000.
- Jones, B.B. & Petkovich, M. (1996): Targeting transcription through nuclear Receptors, *Curr. Pharm. Design*, **2**, 155-168.
- Kamitori, S. & Takusagawa, F. (1994): Multiple binding modes of anticancer drug actinomycin D: X-ray, molecular modeling and spectroscopic studies of d(GAAGCTTC)₂-actinomycin D complexes and its host DNA, *J. Am. Chem. Soc.*, **116**, 4154-4165.
- Karle, J. (1980): Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology, *Int. J. Quant. Chem.*, **7**, 357-367.

- Karle, J. & Hauptmann, H. (1956): A theory of phase determination for the four types on non-centrosymmetric space groups $1P222$, $2P22$, $3P12$, $3P22$, *Acta Cryst.*, **9**, 635-651.
- Klebe, G. & Mietzner, T. (1994): A fast and efficient method to generate biologically relevant conformations, *J. Comput.-Aided Mol. Design*, **8**, 583-606.
- Kubas, M., (2003): Kristallstrukturuntersuchungen an Steroiden und verwandten Verbindungen, *Examensarbeit*.
- Lian, C., Robinson, H., Wang, A.H.-J. (1996): Structure of actinomycin D bound with $(GAAGCTTC)_2$ and $(GATGCTTC)_2$ and its binding to the $(CAG)_n:(CTG)_n$ triplet sequence as determined by NMR analysis, *J. Am. Chem. Soc.*, **118**, 8791-8801.
- Lloyd, A.M., Schena, M., Walbot, V., Davis, R.W. (1994): Epidermal cell fate determination in *Arabidopsis*: patterns defined by a steroid inducible regulator, *Science*, **266**, 435-439.
- Martinez, A., Sparks, C., Hart, C.A., Thompson, J., Jepson, I. (1999): Ecdysone agonist induced transcription in transgenic plants, *The Plant Journal*, **19**(1), 97-106.
- McRee, D. (1999): Xtalview/Xfit – A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol.*, **125**, 156-165.
- Meritt, E.A. & Bacon, D.J. (1997): Raster3D: photorealistic molecular graphics, *Methods in Enzymology*, **277**, 505-524.
- Miller, R., Gallo, S.M., Khalak, H.G., Weeks, C.M. (1994): SnB: Crystal structure determination via shake-and-bake, *J. Appl. Cryst.*, **27**, 613-621.
- Moews, P.C. & Kretsinger, R.H. (1975): Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis, *J. Mol. Biol.*, **91** (2), 201-225.
- Moews, P.C. & Kretsinger, R.H. (1975): Terbium replacement of calcium in carp muscle calcium-binding parvalbumin: an x-ray crystallographic study, *J. Mol. Biol.*, **91** (2), 229-232.
- Otwinowski, Z. & Minor, W. (1997): Processing of X-ray diffraction data collected in oscillation mode, *Meth. Enzymol.* **276**, 307-326.
- Parkinson, G. (1996): New Parameters for the Refinement of Nucleic Acid-Containing Structures, *Acta Cryst.*, **D52**, 57-64.
- Pike, A.C.W., Brzozowski, A.M., Hubbard, R.E., Bonn, T., Thorsell, A.-G., Engström, O., Ljunggren, J., Gustafsson, J.-A., Carlquist, M. (1999): Structure of the ligand binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist, *The EMBO Journal*, **18** (17), 4608-4618.
- Ramachandran, G.N. & Raman, S. (1956): New Method for the structure analysis of noncentrosymmetric crystals, *Curr. Sci. India*, **25**, 348-351.
- Rarey, M., Kramer, B., Lengauer, T., Klebe, G. (1996): A fast flexible docking method using an incremental construction Algorithm, *J. Mol. Biol.*, **261**, 470-489.

- Renaud, J., Bischoff, S.F., Buhl, T., Floersheim, P., Fournier, B., Halleux, C., Kallen, J., Keller, H., Schlaeppli, J.-M., Stark, W. (2003): Estrogen receptor modulators: Identification and Structure-Activity relationships of potent ER α -selective tetrahydroisoquinoline ligands, *J. Med. Chem.*, **46**, 2945-2957.
- Robinson, H., Gao, Y.-G., Yang, X.-L., Sanishvili, R., Joachimiak, A., Wang, A. H.-J. (2001): Crystallographic analysis of a novel complex of actinomycin D bound to the DNA decamer CGATCGATCG, *Biochem.*, **40**, 5587-5592.
- Sasaki, S. (1989): Numerical tables of anomalous scattering factors calculated by the Cromer and Liberman method, *KEK Report*, **88-14**, 1-136.
- Sheldrick, G.M., (2002): Macromolecular phasing with SHELXE, *Z. Krist.*, **217**, 644-650.
- Sheldrick, G.M. & Schneider, T.R. (1997): SHELXL: High resolution Refinement, *Methods in Enzymology*, **277**, 319-343.
- Sheldrick, G.M., Hauptmann, H.A., Weeks, C.M., Miller, R., Usón, I. (1999): Ab initio phasing, In *International Tables for Crystallography Vol. F: Crystallography of Biological Macromolecules*, Edited by M.G. Rossmann & E. Arnold, pp.333-345. Kluwer Academic Publishers, Dordrecht, 2001.
- Shiau, A.D, Barstad, D., Radek, J.T., Meyers, M.J., Nettles, K.W., Katzenellenbogen, B.S., Katzenellenbogen, J.A., Agard, D.A., Greene, G.L. (2002): Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism, *Nature Structural Biology*, **9** (5), 359-364.
- Shinomiya, M., Chu, W., Carlson, R.G., Weaver, R.F., Takusagawa, F. (1995): Structural, physical, and biological characteristics of RNA·DNA binding agent N8-actinomycin D, *Biochem.*, **34**, 8481-8491.
- Simon, R., Igeno, I.-M., Coupland, G. (1996): Activation of floral meristem identity genes in *Arabidopsis*, *Nature*, **384**, 59-62.
- Sybyl Version 6.3. (TRIPOS Inc.; St. Louis, MO), 1995.
- Takusagawa, F., Takusagawa, K.T., Carlson, R.G., Weaver, R.F. (1997): Selectivity of F8-actinomycin D and its anti-leukemia activity, *Bioorg. Med. Chem.*, **5**, 1197-1207.
- Tanenbaum, D.M., Wang, Y., Williams, S.P., Sigler, P.B. (1998): Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains, *Proc. Nat. Acad. Sci. USA*, **95**, 5998-6003.
- Tronrud, D.E. (1992): Conjugate-direction minimization: an improved method for the refinement of macromolecules, *Acta Cryst. A*, **A48**, 912-916.
- Usón, I., Pohl, E., Schneider, T.R., Dauter, Z., Schmidt, A., Fritz, H.-J., Sheldrick, G.M. (1999): 1.7 Å structure of the stabilized REI_v mutant T39K. Application of local NCS restraints, *Acta Cryst. D*, **D55**, 1158-1167.

- Wadkins, R.M. & Jovin, T.M. (1991): Actinomycin D and 7-aminoactinomycin D binding to single-stranded DNA, *Biochem.*, **30** (39), 9469-9478.
- Wadkins, R.M., Jares-Erijman, E.A., Klement, R., Rüdiger, A., Jovin, T.M. (1996): Actinomycin D binding to single-stranded DNA: sequence specificity and hemi-intercalation model from fluorescence and ¹H NMR spectroscopy, *J. Mol. Biol.*, **262**, 53-68.
- Wallace, A.C., Laskowski, Z., Thornton, J.M. (1995): LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions, *Prot. Eng.*, **8**, 127-134.
- Wang, B.-C. (1985): Resolution of the phase ambiguity in macromolecular crystallography, *Methods in Enzymology*, **115**, 90-112.
- Zhang, K.Y.J., Cowtan, K., Main, P. (2001): Phase improvement by iterative density modification, in *International Tables for Crystallography Vol. F: Crystallography of Biological Macromolecules*, Edited by M.G. Rossmann & E. Arnold, pp.333-345. Kluwer Academic Publishers, Dordrecht, 2001.

APPENDIX

Table A. Angles in °.

Angle α							
T102	73,95						
A103	-75,39	A203	114,57	A303	70,82	A403	-64,09
G104	-25,59	G204	-76,55	G304	-71,82	G404	-49,24
BRU105	-167,58	BRU205	-165,13	BRU305	-64,06	BRU405	-177,69
T106	-65,14	T206	116,77			T406	-69,09

Angle β							
T102	-115,76			T302	150,19	T402	168,69
A103	162,45	A203	-175,86	A303	-136,08	A403	-161,98
G104	158,4	G204	173,23	G304	-179	G404	134,6
BRU105	145,57	BRU205	99,9	BRU305	145,92	BRU405	127,45
T106	171,39	T206	174,88			T406	164,39

Angle γ							
T102	-118,81			T302	20,31	T402	24,6
A103	45,96	A203	76	A303	53,93	A403	53,2
G104	26,66	G204	17,22	G304	40,46	G404	39,92
BRU105	170,9	BRU205	159,27	BRU305	44,07	BRU405	-165,89
T106	46,33	T206	98,16			T406	57,92

Angle δ							
T102	154,41			T302	142,54	T402	139,3
A103	92,59	A203	150,51	A303	147,68	A403	126,26
G104	138,31	G204	136,83	G304	129,48	G404	114,17
BRU105	90,75	BRU205	85,11	BRU305	88,15	BRU405	111,36
T106	86,39					T406	82,25

Angle ε							
T101	-177,03						
T102	-144,56			T302	-156,99	T402	157,31
A103	-132,53	A203	-141,64	A303	-178,02	A403	-95,93
G104	-159,46	G204	-140,73	G304	-103,36	G404	-174,85
BRU105	-179,23	BRU205	-149,51	BRU305	-153,84	BRU405	-178,89

Angle ζ							
T102	67,27						
A103	-96,18			A303	159,01	A403	-88,57
G104	-83,65	G204	-90,84	G304	-87,86	G404	-141,58
BRU105	-76,98	BRU205	-71,9	BRU305	-75,56	BRU405	-74,72
T106	-91,16	T206	-92,32			T406	-91,16

Angle χ							
T101	-167						
T102	-151,35			T302	-146,53	T402	-102,81
A103	-127,45	A203	-100,27	A303	-98,85	A403	-109,64
G104	-81,58	G204	-73,54	G304	-95,23	G404	-71,34
BRU105		BRU205		BRU305		BRU405	
T106	-109,94	T206	-164,41			T406	-121,41

Danksagung

An erster Stelle möchte ich meinem Doktorvater, Prof. G.M. Sheldrick, Ph.D., für die interessante Themenstellung, die ausgezeichneten Arbeitsbedingungen und die stete Hilfs- und Diskussionsbereitschaft ganz herzlich danken.

Mein Dank gilt auch Prof. Dr. Isabel Usón für die Betreuung der kristallographischen Projekte, sowie Ihr Interesse an meine Arbeit.

Besonderer Dank gebührt Dr. Martina Schäfer, die mir ermöglichte Einblicke in die pharmazeutische Industrie zu erlangen, sowie neue Methoden zu erlernen. Darüber hinaus für die stetige Unterstützung und Interesse an meine Arbeit. Und nicht zuletzt sei Ihr und Alex, sowie Yoda und Yedi, für die herzlich Aufnahme in Ihr Heim gedankt.

Für die Einarbeitung und die Tipps zum *Molecular Modeling* danke ich herzlich Dr. Ursula Egner, Dr. Judith Günther, Dr. Hans Briem und für die sehr freundliche Atmosphäre alle Mitglieder der *Structural Biology* sowie der *Computational Chemistry* Abteilungen der Schering AG.

Für die gute Zusammenarbeit im Actinomycin Projekt, sowie die Bereitstellung des 7-AAMD und der Oligomere, sei Prof. T. Jovin, Herrn Reinhard Klement, Frau Jares-Erijman, sowie Frau Gudrun Heim gedankt.

Für die Bereitstellung des H-(L-Tyr-D-Tyr)₄-Lys-OH Peptides, sowie die ausgezeichnete Zusammenarbeit, sei Prof. U. Diederichsen und Dipl. Chem. Andrea Küsel gedankt.

Gedankt sei auch Prof. Tietze und Dipl. Chem. Carsten Vock für die stete Hilfs- und Diskussionsbereitschaft.

Ich danke allen Mitgliedern unserer Abteilung für die angenehme Arbeitsatmosphäre und speziell Verena König und Christine Schlicker für die lustige Zeit in unserem Arbeitszimmer. Mein herzlicher Dank gilt ausserdem Fabio Dall'Antonia, Verena König, Ilka Müller und Thomas Pape für die stete Hilfsbereitschaft und mentale Unterstützung in Krisensituationen, auch über die Arbeit hinaus. Ilka Müller sei besonders gedankt für die Messung des d,l-Peptides in Hamburg.

Für das Korrekturlesen danke ich Dr. Ilka Müller, Dr. Fabio Dall'Antonia, Dipl. Chem. Christine Schlicker, Dipl. Biol. Tine Kragh Nielsen, Dr. Martina Schäfer und Prof. Dr. Isabel Usón.

Gedankt sei auch Molly und Jimmy für deren Hilfe beim Seeding.

Am herzlichsten aber sei meiner Familie gedankt für Ihre stetige Unterstützung und Liebe während der gesamten Studienzzeit.

Und nicht zuletzt danke ich Markus für die schöne Zeit und dafür, dass ich auf ihn zählen kann!

Lebenslauf

Persönliche Daten

Name	Eftichia Alexopoulos
Geburtsdatum	13. Dezember 1978
Geburtsort	Athen, Griechenland
Staatsangehörigkeit	deutsch und griechisch
Familienstand	ledig

Studium und schulische Ausbildung

Juli 1996	Griechisches Abitur, 1. Lyzeum Ano Ilioupoli, Athen, Griechenland
Okt 1996	Diplomstudiengang Chemie, Friedrich-Wilhelm-Universität Bonn
Sep 1998	Vordiplom Chemie, Friedrich-Wilhelm-Universität Bonn
Okt 1998	Immatrikulation an der Georg-August-Universität Göttingen im Studienfach Chemie
November 2001	Diplomarbeit am Lehrstuhl für Strukturchemie der Georg-August-Universität Göttingen unter der Leitung von Prof. George M. Sheldrick, PhD: <i>Röntgenstrukturuntersuchungen an Vinamidinmetallkomplexen sowie Kristallisation eines 7-Aminoactinomycin-ssDNA Komplexes</i> Abschluss: Diplom-Chemikerin

Promotion

Nov 2001- April 2004	Dissertation am Lehrstuhl für Strukturchemie der Georg-August-Universität Göttingen im Arbeitskreis von Prof. G.M. Sheldrick, PhD: <i>Crystallographic and modeling studies of intermolecular interactions of biological interest</i>
----------------------	--