UNIVERSITY OF SOUTH BOHEMIA

IN ČESKÉ BUDĚJOVICE

FACULTY OF BIOLOGICAL SCIENCES

MASTER'S THESIS

STRUCTURAL STUDIES OF THE HALOALKANE DEHALOGENASE MUTANT (DHAA12) FROM *Rhodococcus rhodochrous*

Bc. Jiří Emmer, DiS.

Adviser: Mgr. Ivana Kutá Smatanová, Ph.D.

Consultant: RNDr. Jiří Brynda, CSc.

České Budějovice, 2007

Master's Thesis

Emmer, J., 2007. Structural studies of the haloalkane dehalogenase mutant (DhaA12) from *Rhodococcus rhodochrous*, Mgr. Thesis, in English. – 39 p., Faculty of Biological Sciences, University of South Bohemia in České Budějovice, Czech Republic.

Annotation:
Common crystallization procedures, X-ray diffraction method and crystallographic software to determine and refine the structure of haloalkane dehalogenase enzyme were used in this thesis.

I declare, that in accordance with § 47b of the law No. 111/1998 Sb. in valid version I agree with publication of my Master's Thesis in unabridged form the electronic way in public accessible part of the STAG database run by the University of South Bohemia in České Budějovice on its website.

I declare that I elaborated the master's thesis alone only with help of the cited literature.

In České Budějovice 7th May 2007

……..

Acknowledgement

At this place I want to give great thanks to my advisor Ivana Kutá Smatanová for her patience and qualified leading during elaborating of my master's thesis. I'm also very grateful to my consultant Jiří Brynda, for his invaluable help with measuring and processing of data. I also want to thank my parents for great moral and financial support not only during the time of my studies and also my friends for help with carrying of the burden of academic studies.

AIM OF PROJECT

The aim of this project is:

a) crystallization of the mutant of haloalkane dehalogenase enzyme
b) testing of protein crystals
c) solving of protein structure

This project is supported by the grant LC06010 of the Ministry of Education of the Czech Republic.

CONTENT

I. THEORY

Proteins play an important role in all fundamental processes of life. Various methods such as electron microscopy, biochemical and biophysical spectroscopic methods, electron crystallography etc. have been used to study proteins, yet to determine their structure, in many cases, remains problematic. The fact that macromolecular crystals are quite unstable, extremely fragile and sensitive to external conditions, they have to be kept in a solvent-saturated environment, otherwise dehydration will lead to crystal cracking and destruction. These reasons are responsible for difficult crystal growing and solving their structure.

How to get the structure?

To solve a structure of macromolecule, several steps have to be realized a priori.

1. To obtain protein monocrystal,
2. To test monocrystals (protein crystals × salt crystals),
3. To carry on diffraction experiment,
4. To process collection data, integrate all diffraction pictures and convert them into a digital map,
5. To scale data – using program that determines intensities of peaks and convert them to structure factors,
6. To solve the phase problem. The most common and easiest way is molecular replacement. We use a program that can do molecular replacement on the scaled raw data and build a PDB file – a map file of every single atom with the information about the amino acid, where each atom is situated.
7. To do structure refinement. PDB file serves as input for next main programs, where the final part of this procedure – fitting of amino acids of the model into the electron density map and refinement – is done.
8. Validation of protein structure.

Why do we need a crystal?

The answer is simple: as a Fig. 1. indicates, a diffraction of a single molecule or a very little crystal is not sharp enough to show the diffraction spots on the detector. And molecules in a solution are not ordered well enough to provide diffractions all in one direction, also it is also

not possible to obtains the diffraction spots. We need many protein molecules ordered in one direction and the ideal way to achieve this is to grow the crystal.

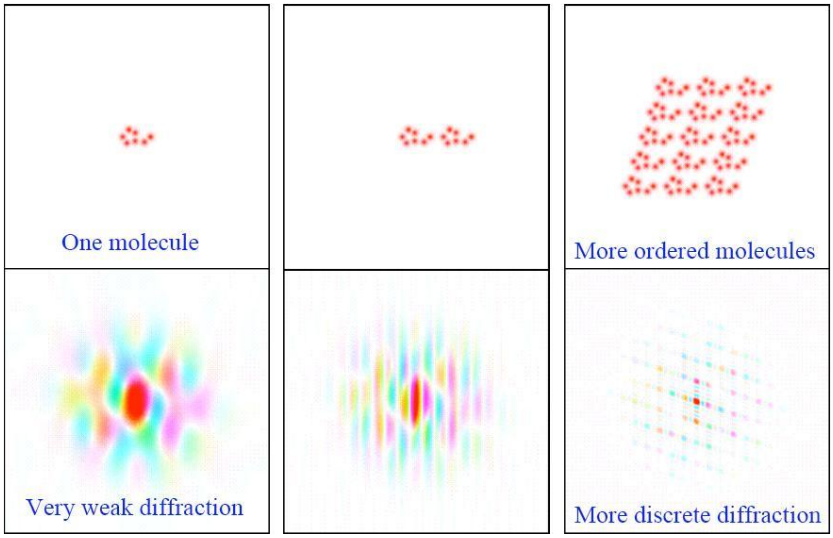

Figure 1. A reason why we need a crystal [8]

1. Crystallization of Protein [1, 2, 3, 4]

In order to obtain a structure of a protein using the X-ray diffraction method, we need to get the crystal first. Comparing to a fair amount of inorganic matters, proteins need very specific conditions to make their crystal's growth.

Protein crystallization is mainly the trial-and-error procedure in which the protein is slowly precipitated from its solution. There are several methods to help us determine the crystallization conditions, but so far no theoretical method how to determine them without testing has been developed.

A success of crystallization procedure depends on several factors:[7]:

1. purity of a protein. We need an extremely pure protein to achieve the crystallization, because otherwise the impurities would cause distortion of the initial crystal cell. Though in some cases they can actually help the protein to crystallize, this is for the cost of a lower resolution of the measured data
2. a suitable precipitant (usually a chemical soluble in water). Often used precipitants are inorganic and organic compounds such as ammonium sulphate, sodium chloride, 2-methyl-2,4-pentanediol (MPD), etc. or macromolecular compounds - polyethylene glycol (PEG) of various molecular weights. Precipitant is added only to such

concentration that will not cause the protein to form insoluble precipitates. Hydrophobic membrane proteins require addition of a detergent.

3. various conditions influencing crystal growth like temperature, pH, presence of other solution components. We are trying to bring the protein to supersaturation, when the concentration of a protein is brought above its solubility limit. At this point, the protein begins to aggregate.

Crystallization procedure can be described in following steps:

1. First step of an actual crystal growth is nucleation, when nuclei, small amorphous stable protein aggregates are created. Because aggregation occurs more easily than crystal growth, we must ensure that the protein is not brought to supersaturation too quickly otherwise only precipitate is appeared.

2. Second step is actual growth. As the crystals grow, the concentration of protein in solution is decreasing as well as supersaturation. It is necessary to achieve slow growth in order to produce large and well-ordered crystals. To achieve a slow growth usually temperature or protein concentration are lowered.

Relation between concentrations of protein and crystallizing agent (most often the precipitant) can be demonstrated using phase diagram [18].

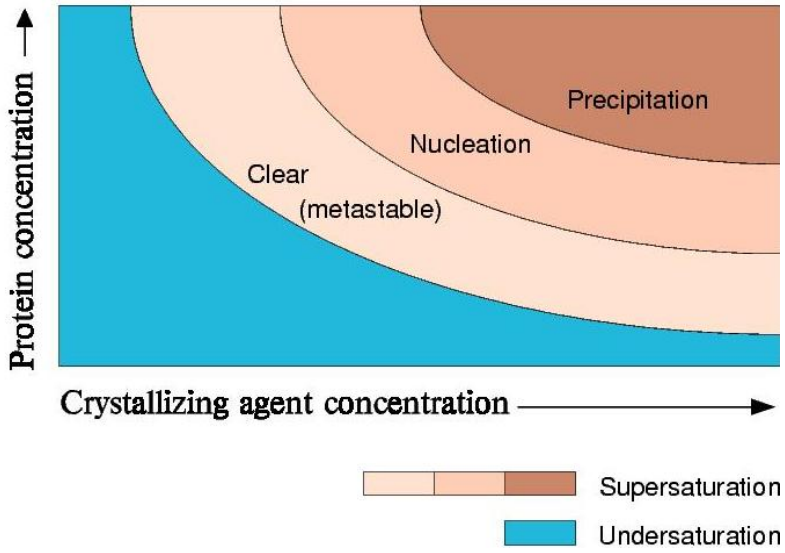

Figure 2. Phase diagram

To achieve crystal growth, supersaturation has to be reduced to a lower level as a high supersaturation (precipitation area of the diagram) would result in the formation of many

small nuclei and/or small crystals. The ideal way how to do it is to reach such protein and precipitant concentrations, which lie in the nucleation area of the phase diagram. The crystals should then grow slowly to reach a maximum degree of order into their structure. This happens in the metastable region, where no nucleation is possible, only crystal growth. The easiest way to change the supersaturation is to change temperature. However there is a problem with phase diagram, where is no easy to determine lines, which separate particular regions in the diagram. This approach is mainly theoretical, but still useful.

There are also other ways to achieve the protein precipitation. One of these is a "salting out" experiment that consists in increasing effective concentration of a protein by adding the salt like ammonium sulfate.
The opposite method, "salting in" is used when the protein is poorly soluble in pure water and adding of a small amount of salt improves its solubility. When the salt is removed, the protein precipitates.

1.1 Crystallization Experiments

The procedure of protein crystallization can be divided in two steps:

1. Testing various crystallization conditions. Usually commercial kits developed for this purpose are used. If a similar protein has already been crystallized, we also test conditions used for this protein.
2. After testing is done, the second step is optimization crystallization conditions, which showed to be at least partially suitable. An optimization is done by systematically varying conditions like pH, concentration of precipitant, protein concentration or other reagents and also by changing temperature. Lowering temperature means slowing a crystal growth. The most common buffers are intended to be effective in the pH range of 6.0 to 8.0, because the most physiologically important reactions occur near to neutral pH.

1.2 Crystallization Techniques

There are many methods how to crystallize biological macromolecules. Aim of all of them is to bring the solution of macromolecules to a supersaturation state. In this part only one will be described, Sitting drop vapor diffusion technique, which we used for crystallization experiments is described below. Other techniques are described in detail in [36] and [1, 2, 3, 4, 5, 7]

Sitting Drop Vapor Diffusion Method [7]

The most frequently used crystallization method is a vapor diffusion technique. The difference in concentration between a drop and a reservoir drives the system toward equilibrium by diffusion through the vapor phase. The protein then becomes supersaturated and crystals start to form when the drop and reservoir are at/close to equilibrium.

In this version of vapor diffusion technique the drop really sits in a small depression a little higher than the precipitant solution is situated. The chamber is usually sealed with a duct tape; we use a crystal clear tape approved for usage by US mail service, which is very transparent and this avoids problems with viewing the result.

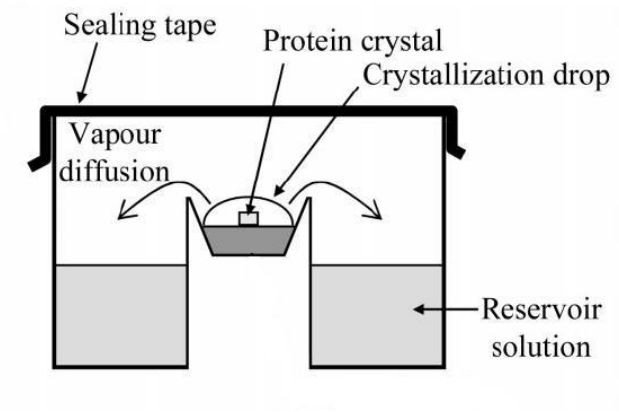

Figure 3. Sitting drop method

2. Testing of Crystals

Several methods are available to test character of crystals:

- Crush test – if we have enough crystals, we can simply crush the crystal with a needle. If we manage to damage it easily, it is a protein, otherwise it is a salt.
- Dehydration test – if enough crystals are grown, we remove the cover and let crystals dry out. If the crystal shrinks, it is a protein. If a nice crystal remains, it is salt.
- Izit (Hampton Research, CA, USA) or methylene green dye-binding test – we let the dye soaking into crystal. If the dye penetrates into a crystal and crystal becomes colored, it is a protein, but the dye cannot penetrate the ion interactions in the salt, so the crystal remains clear. These colored protein crystals can be used for diffraction experiment, but in any cases a dye would cause extra diffractions and it would be difficult, even impossible to distinguish between them.

- SDS PAGE electrophoresis – we dissolve crystals again and run electrophoresis. A protein will remain in a position corresponding to its molecular weight, while salt will run out.
- X-ray diffraction – in the case of salt we get a diffraction pattern typical for salt (scarce strong diffraction), but if the crystal is a protein, we get the diffraction pattern typical for a protein (many spots in several concentric circles). This is the only advisable way of testing crystals, if we have just small amount of them.

3. Diffraction Experiment

After obtaining protein monocrystals, diffraction measurement is performed. This is done on a synchrotron or protein diffractometer [8]. Protein diffractometer is a measuring device composed of the source of X-rays, a fixation device placed on a goniometer and image plate detector. The primary X-ray beam is monochromated by either crystal monochromators or focusing mirrors. The beam passes through the crystal mounted on a pin of a goniometer head, which is being cryocooled, usually by a stream of low temperature nitrogen. This is an important safety measure, because before this was brought into practice, many protein crystals were damaged soon by the focused high-energy X-ray radiation and usually more crystals were needed to record the whole angle range that was needed. The goniometer, where the head with the crystal is mounted, allows positioning of crystal in different orientations in the beam. The diffracted X-rays are recorded using an image plate detector. The principle of this device is shown at Fig. 4.

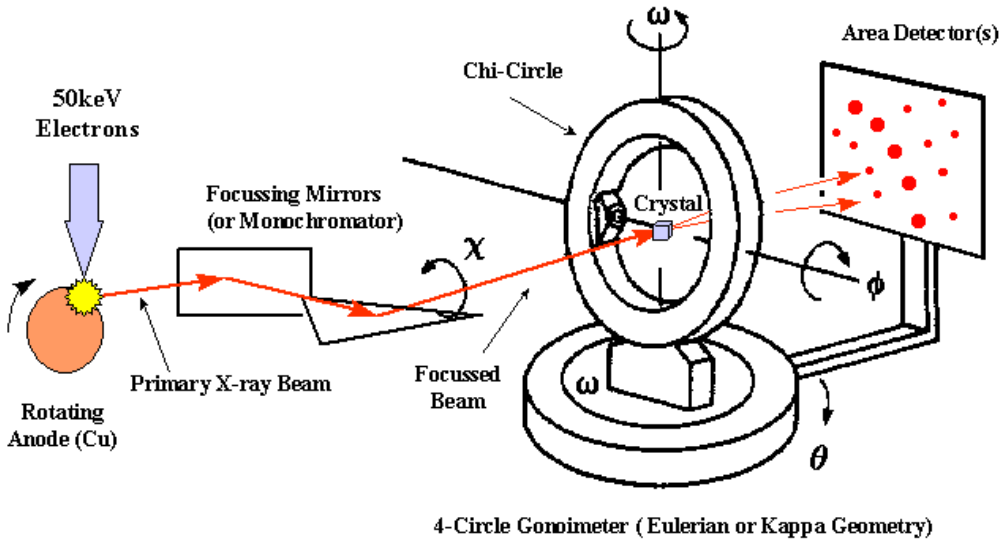

Figure 4. Protein diffractometer [8]

The process of measurement is following:

- we fish out the crystal, catching it in a little loop,
- crystal is then cooled down, usually putting directly into a liquid nitrogen,
- crystal is fixed into the goniometer while the cooling device is still running,
- the X-ray source is activated,
- then we or the automated device rotate with the crystal in the goniometer,
- each turn of the certain little angle (in most cases 1°) is recorded on an image plate and brought to the computer. We can set various time values for exposition of crystal in one position, depending on the state, temperature and quality of the crystal.

4. Processing of Data Collection [26]

The main program for data processing is "MOSFLM" [10]. This program is part of the CCP4 Program Suite [6], on the other side this program is also distributed standalone, even in the form of precompiled program files.

The task of this program is to produce from a set of diffraction images a set of indices (*hkl*) with their associated intensities, estimates of their uncertainties and the estimate of the unit-cell parameters. This process can be divided into three steps:

1. autoindexing - it determines crystal orientation, unit-cell parameters and from them the possible space group,
2. post-refinement – it is to refine the initial estimate of unit-cell parameters and also crystal mosaicity,
3. integration of images - it consists of predicting the positions of the Bragg reflections on each image and obtaining an estimate of the intensity of each reflection and its uncertainty. There are two main ways of integration – two-dimensional and three-dimensional integration. When using two-dimensional integration, the intensities of the different components of a partially recorded reflection (on different but adjacent images) are evaluated independently by two-dimensional profile fitting and only summed to give the total intensity when the data are merged and scaled. By contrast, when using three-dimensional integration the different components are assembled by the integration software and a three-dimensional profile is used to evaluate the total intensity. MOSFLM uses two-dimensional integration. The intensity is determined by counting pixels containing the spot and then subtracting the X-ray background, which

is taken in the close surrounding, because logically it cannot be measured directly under the spot. Then the reflection lists is generated into an MTZ file, which includes the reflection map, intensity for each reflection and a standard deviation of each reflection.

The MOSFLM program is designed to facilitate processing of rotation data collected on image plate, CCD or film.

5. Scaling of Data [13, 30]

Data scaling is done using program named "SCALA". SCALA is an integral part of the CCP4 Program Suite [6] that can be found in the "Data reduction" section.

The program scales together multiple observations of reflections and merges multiple observations into an average intensity. The diffraction intensities measured by integrating spots recorded on an area detector are not all on the same scale because they are affected by a number of physical factors from an experiment most of them are difficult to measure directly. The process of 'data reduction' uses the redundancy of multiple measurements of symmetry-related reflections to put all observations on a common scale by fitting a scaling model that reflects the experiment. This process produces a data set that is internally consistent, within errors of the model, though not necessarily correct on an absolute scale. Analysis of agreement between equivalent reflections after scaling gives estimates of a quality of data and also highlights parts of the data that agree poorly with the rest. This allows decisions to be made about whether parts of the data should be rejected.

Measured diffraction intensities are affected by various physical factors, which can be divided into two groups: 1. factors related to the incident X-ray beam and 2. factors related to the crystal and diffracted X-rays.
Factors related to the incident X-ray beam are e. g. variations in the rotation rate, rapid fluctuations in incident beam intensity or errors in synchronization of the shutter, which cause systematic errors that are difficult or impossible either to detect or to model. From this reason these problems should be monitored. Among the factors related to the crystal and diffracted X-rays we can name absorption in the secondary beam direction, which is serious at long wavelengths, but the most difficult systematic error is radiation damage, since radiation causes the structure to change with time, which means that different reflections change at

different rates. Extrapolation to zero time requires many observations of each reflection well spaced out in time and this is not generally possible in radiation-sensitive cases.

After scaling, the intensities are analyzed: 1. to set the real resolution of the data set, 2. to detect bad regions (e.g. bad images), 3. to analyze radiation damage and 4. to assess the overall quality of the data set. The significance of any anomalous signal may be assessed by probability and correlation analysis. Knowledge of the Laue group and point-group symmetries is required for the scaling and merging of intensities: the possible symmetry of the diffraction pattern may be determined from scores such as correlation coefficients between observations which might be symmetry-related.

If we are working with the CCP4 suite, we use the program “TRUNCATE”. This program reads a reflection data file of averaged intensities created by SCALA and outputs an MTZ reflection data file containing structure factor F and DeltaFanom values. The input intensities are assumed to follow a normal distribution with standard deviations, i.e. negative observations have to be preserved. The truncation procedure is based on Bayesian statistics. The structure factors (F's) are calculated using the prior knowledge of Wilson's distributions for acentric or centric data and the mean intensity and standard deviation values. The F's outputs are all positive and follow Wilson's distribution. The truncation procedure has little effect on reflections larger than three standard deviations but should give significantly better values for the weak data than those obtained by merely taking the square root of the intensities and setting negative intensities to zero. Reflections of less than minus four standard deviations are rejected.

For more information see [13, 30].

6. Solving of Phase Problem [26]

In an ideal diffraction experiment with a strictly parallel monochromatic X-ray beam and a perfect crystal, the Ewald sphere is an infinitely thin spherical shell and the reciprocal lattice points are infinitely small, so that each Bragg reflection occurs only at a precisely defined φ value (when the reciprocal-lattice point intersects the Ewald sphere).

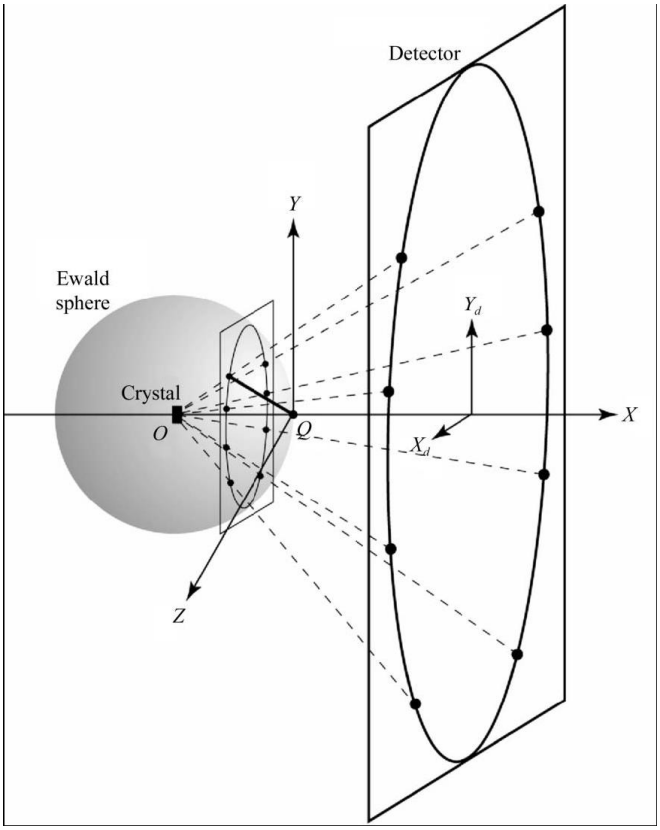

Figure 5. Ewald sphere construction

The Ewald sphere construction. The X-ray beam is along the X axis and the Z axis is the rotation axis. The origin of the reciprocal lattice lies at the point that the X-ray beam exits the Ewald sphere (Q) and the crystal is located at the centre of the sphere (O). The crystal is oriented so that the principal zone axis lies along the X axis. The reciprocal-lattice plane h = 1 is shown. Each diffraction spot on the detector can be mapped back to the equivalent scattering vector in reciprocal space. One such scattering vector is shown as a bold line from the reciprocal-lattice origin (Q) to the surface of the Ewald sphere.[26]

In reality, the divergence and wavelength dispersion of the X-ray beam result in an Ewald sphere with a finite thickness, while crystal mosaicity (and intrinsic variation in unit-cell parameters between different mosaic blocks) means that each reciprocal lattice point has a finite size. The combination of these effects means that each reflection has a finite reflecting range (or width) in φ. In most cases, the reflecting range is determined primarily by the mosaic spread of the crystal (particularly for cryocooled crystals that often have a mosaic spread greater than 0.3°) and by geometric factors that give rise to the Lorentz correction. This correction arises because for reciprocal-lattice points of any given size the φ rotation required for them to pass entirely through the Ewald sphere is proportional to their distance from the rotation axis. The practical result of this finite reflecting range is that reflections may be recorded only on a single image (fully recorded) or on two or more images (partially recorded) depending on the mosaic spread and the rotation angle per image (Fig. 6).

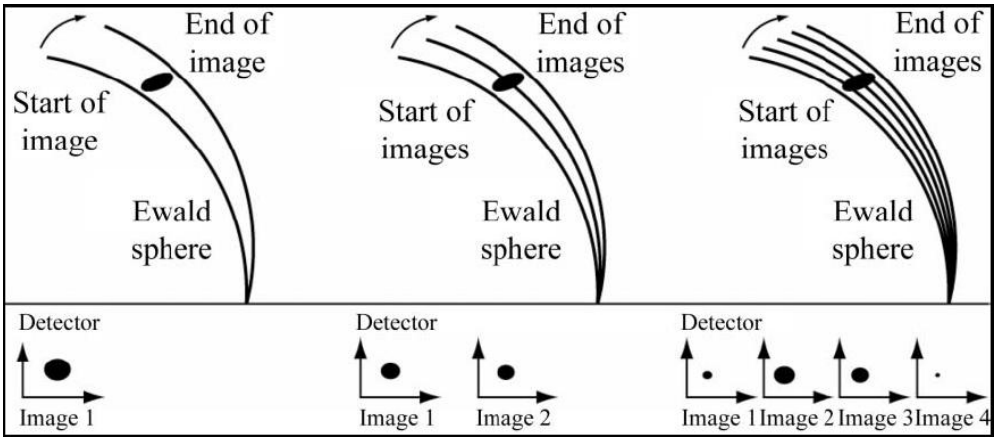

Figure 6. Schematic representation of the effect of the finite size of reciprocal lattice points on the diffraction images. [26]

The black ellipse represents a reciprocal-lattice point of finite size that results from the combination of crystal mosaicity, beam divergence, wavelength dispersion and variability in unit-cell parameters. The arcs represent the positions of the Ewald sphere at the beginning and end of one or more images. When the oscillation angle is large compared with the angular width of the reciprocal-lattice point (left), the entire reciprocal-lattice point lies between the two arcs and all the intensity is recorded on a single image (a fully recorded reflection). If the oscillation angle per image is halved (centre), the total intensity is distributed over two images (partially recorded reflections). If the oscillation angle is significantly less than the reflecting range, the intensity is distributed over several images (right), resulting in fine φ-slicing. Figure after Elspeth Garman.

Here we see that we need to determine the phase exactly in order to get the correct electron density map. There are several methods how to solve this phase problem.

6.1. Molecular replacement (MR) [33]

If a related or homologous (more than 50% sequence identity) protein structure is already known then the method of Molecular Replacement (MR) may be used. The idea is to find the rotation and translation, which position the model structure in the unit cell as to give the highest correlation between experimental diffraction measurements and those calculated from the model. More information about this method can be found in [36] or [7, 11].

6.2. Multiple isomorphous replacement (MIR, MIRAS) [32, 33]

The most common method of ab initio phase determination for proteins is Multiple Isomorphous Replacement (MIR) - introduced 40 years ago. MIR involves introducing heavy atoms into protein structure. Heavy-atom derivatives are prepared by soaking the native protein crystals in a buffer containing heavy-atom compound or by cocrystallization. The goal is to obtain derivatives crystals, in which heavy atoms bind specifically and consistently to each protein molecule in the unit cell. After soaking, the positions of the heavy atoms are determined using difference Pattersons. For this step to be successful it is important that only

a few heavy atoms bind in each asymmetric unit. Once the initial heavy atom locations have been found, the coordinates, occupancy and temperature factors of each heavy atom are refined. At least two isomorphous derivatives are needed for successful MIR refinement, and for MIRAS phasing (MIR and Anomalous Scattering), one isomorphous derivative plus anomalous scattering data are needed. In practice, data from several derivatives are combined for the refinement of heavy atom parameters and for the calculation of an MIR or MIRAS phased Fourier map that is suitable for building the initial structure. The native and heavy-atom derivative crystals should be isomorphous - crystal form and unit cell dimensions should be unchanged. The structure factors of heavy-atom derivative (ph) and native protein (p) and heavy atom (h) have the following relationship:

$$\mathbf{F_p = F_{ph} - F_h} \qquad (6.1)$$

If the magnitudes $|F_{ph}|$ and $|F_p|$ are measured experimentally, $|F_h|$ and phase angle are obtained from the heavy-atom model. Only one heavy-atom derivative (single isomorphous replacement - SIR) gives two possible solutions for the phase of F_p. One of these is the true solution.

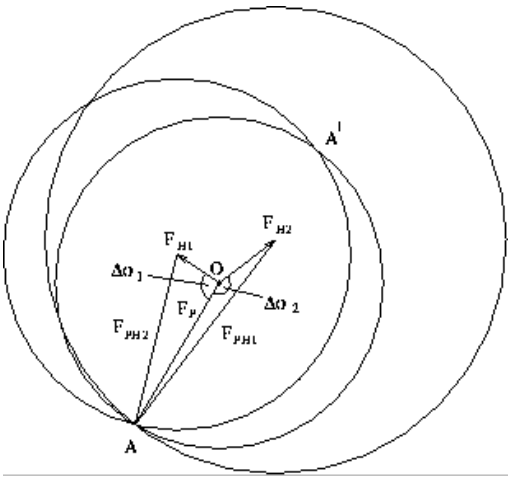

Figure 7. Using second heavy-atom derivative (ph2) with a different position, a unique solution can be obtained as shown in a geometric representation. [33]

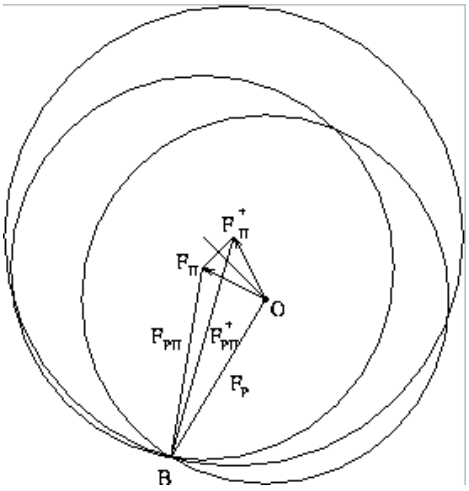

Figure 8. A similar vector construction also can be produced by anomalous dispersion. [33]

6.3. Multiple wavelength anomalous scattering [31]

Generally, in protein X-ray diffraction experiments is assumed that interaction of X-rays with electrons gives rise to elastic scattering of the incident radiation. However, when the energy of the incident X-rays is closed to an electronic transition in a bound atomic orbital, resonance resulting in anomalous scattering occurs. . Therefore, the total scattering factor is usually written

$$\mathbf{f = f_0 + f' + if''} \qquad (6.2)$$

where f is the total scattering, f_0 is the normal or Thomson scattering and f′ and if ′′ are the real and imaginary components of the anomalous scattering. The magnitude of the anomalous scattering is essentially independent of the scattering angle, but does depend strikingly on the energy of the incident radiation. The intensity of the normal scattering, on the other hand, falls off with increasing scattering angle. Absorption of an X-ray photon provides the energy for an electronic transition, which perturbs the amplitude and phase of the diffracted X-rays and gives rise to anomalous scattering. The characteristic energy for a specific transition is referred to an absorption edge and can be easily characterized by X-ray absorption spectroscopy. Absorption edges are named with respect to the atomic orbital, from which they originate. Thus, electronic transitions that originate from 1s orbitals are referred as K edges, from 2p orbitals - L edges and from 3d orbitals - M edges. For the atoms C, N, O and S, the magnitude of the anomalous signal is negligible and usually ignored. However, anomalous scattering from sulfur has been successfully used to obtain phase information in a number of cases. Anomalous scattering effects are more easily detectable for heavier atoms. For non-centrosymmetric structures, as the imaginary component (f ′′) of the anomalous scattering is

90° out of phase to that of the normal scattering, Friedel's law, $\left[I_{(hkl)} = I_{\overline{(hkl)}}\right]$ no longer holds. The resultant intensity differences in the Friedel pairs, now referred to as Bijvoet differences, can be used in a somewhat similar way to MIR to obtain phase information.

6.4. Direct methods [34]

Direct methods can be used if very accurate data are available.
There are two major processes for recovering the phases using the data obtained by regular equipment. One is the direct method, which estimates the initial phases and expanding phases using a triple relation. (A trio of reflections in which the intensity and phase of one reflection can be explained by the other two has a triple relation.) A number of initial phases are tested and selected by this method.

The other is the Patterson method, which directly determines the positions of heavy atoms. The Patterson function gives a large value in a position that corresponds to interatomic vectors. This method can be applied only when the crystal contains heavy atoms or when a significant fraction of the structure is already known.

For molecules whose crystals provide reflections in the sub-angstrom range, it is possible to determine phases by brute force methods, testing a series of phase values until spherical structures are observed in the resultant electron density map. This works because atoms have a characteristic structure when viewed in the sub-angstrom range. The technique is limited by processing power and data quality. For practical purposes, it is limited to "small molecules" since they consistently provide high-quality diffraction with very few reflections.

Next direct method is a reference-beam diffraction data-collection technique [35]. Using this technique it is possible to directly measure a large number of relative phases of Bragg reflections on an area detector in a typical protein crystallography experiment. The technique, being developed at Cornell, incorporates the principle of three-beam diffraction into the most common method of data collection, i.e., the oscillating-crystal method, and allows recordings of many phase-sensitive three-beam interference profiles simultaneously. Recent advances include a dedicated five-circle κ diffractometer and new data acquisition and analysis

algorithms. Experimental results on a protein crystal are presented [35] and the strategies of using the measured phases for solving crystal structures are discussed [35].

7. Fitting of Amino acids and Structure Refinement

7.1. Fitting of Amino acid

The last main part of structure solving is fitting of amino acids and following refinement. The actual fitting can be done using program "XFIT". The program is the part of the suite XTalView [14], which is not the integral part of the CCP4 Program Suite, although cooperation of these two packages is possible. XFIT is a large and complex program of that main purpose is to draw electron maps and fit amino acid residues of the model into them. The model is usually in PDB format and replacing, adding or deleting of amino acids can edit the model itself. Some of the more notable features are:

- Model-fitting to electron density.
- A built-in FFT to calculate density from structure factors.
- Simultaneous refinement and fitting - the geometry anneals as the model is fit and the allowed angles inside the amino acids are preserved.

7.2. Structure Refinement

The "REFMAC5"is a program used for refinement of the model of a protein (or another macromolecule). It is integral part of the CCP4 Program Suite [6]. The program has two modes: REVIEW that only checks and updates the input model so the geometric restraints like the disulphide (and other covalent) bonds and cis-peptides can be properly set up, and the standard REFINE mode, where the program minimizes the coordinate parameters to satisfy either a Maximum Likelihood or Least Squares residual. The output is then the PDB file with the model and the MTZ file with refined electron density maps.

More information about the REFMAC5 program is available at [12].

7.2.1. The traditional crystallographic R-factor [15]

The purpose of a crystallographic R-factor is to show using a single number how a complex model composed of thousands of atoms agrees with a large set of individual diffraction observations. Typically, thousands of atoms determine thousands of computed structure factors, which are then compared to the observed structure factors to yield an R-factor.

The traditional crystallographic R-factor is defined as

$$R = \frac{\sum |F(obs) - F(calc)|}{\sum F(obs)} \qquad (7.1)$$

This R-factor is a measure of the level of disagreement between the observed structure factors (F_{obs}) and calculated structure factors (F_{calc}) from our model. It is usually reported in %, i.e. an R-factor of 0.18 is reported as 18%.

This R-factor is not very sensitive to errors, especially in case of:

- low-resolution structures with a low observations-to-parameters ratios
- overall good structures with few localized errors
- structures with distorted geometry

7.2.2. The Free R factor [15]

The free R-factor shows greatly reduced coupling to the target function used during minimization (the residual) than the R-factor. The reason is that it compares the atomic model with a small part of data, which has been omitted from the refinement that means it shows the disagreement of the model with an original data, which is not affected by refinement.

For this reason already after scaling is performed a random subset (5-10%) of the dataset is set aside and labeled the free or test set. The remaining 90-95% of the dataset (working set) is used to form the target function for refinement and to compute the traditional crystallographic R-factor.

In contrast, the R_{free} is the R-factor calculated for the test set alone. The R_{free} is commonly 2-8% higher than the regular R-factor.

8. Properties of the DhaA haloalkane dehalogenase

The DhaA protein is an inducible hydrolytic haloalkane dehalogenase and this enzyme is naturally found in the *Rhodococcus rhodochrous* bacteria. Natural environment of mentioned bacterium is soil polluted with haloalkanes and the bacterium is capable of utilizing 1-chlorobutane and several other haloalkanes as the sole carbon and energy source. This feature is important, because these bacteria can perform biodegradation of dangerous chemicals, which haloalkanes without any doubt are.

In general, degradation can be based on one of four processes: 1. the use of a halogenated compound as a carbon source and oxidation able substrate with either oxygen or nitrate as an electron acceptor; 2. fermentative metabolism, in which a dehalogenated intermediate serves as electron acceptor; 3. the use of a halogenated compound as an electron acceptor with liberation of halide; and 4. co-metabolic transformation and halide release linked to any metabolite process. The last process arises through the lack of specificity of an enzyme with another physiological function. The key reaction during microbial degradation of halogenated compounds is the actual dehalogenation. During this step, hydrogen or a hydroxyl group replaces the halogen substituent, which is usually responsible for the toxic and xenobiotic character of the compound. Halogen removal reduces both recalcitrance to biodegradation and the risk of forming toxic intermediates during subsequent metabolic steps. The latter may occur during oxidative conversion, where the presence of halogen substituents may lead to the production of acylhalides or 2-haloaldehydes, which are reactive products owing to their electrophilicity and may cause cellular damage [27].

The DhaA protein is a member of a haloalkane dehalogenase family, type 2 subfamily. Its purpose is to catalyze hydrolytic cleavage of carbon-halogen bonds in halogenated aliphatic compounds, which leads to formation of the corresponding primary alcohols, halide ions and protons. In brief we can describe the function with the following formula:

$$\text{1-haloalkane} + H_2O = \text{primary alcohol} + \text{halide}.$$

The enzyme expresses halogenase activity against 1-chloroalkanes of chain length C_3 to C_{10}, and also shows a very weak activity with 1,2-dichloroethane.

8.1. Native DhaA haloalkane dehalogenase [21, 22]

Native dehalogenase consists of 293 amino acids. The enzyme has 3 active sites such as a nucleophile at the leucine Leu106, a proton donor at asparagine Asn130 and a proton acceptor at asparagine Asn272 [23]. All these residues are situated on the surface of the molecule, as is shown on the Fig. 10.

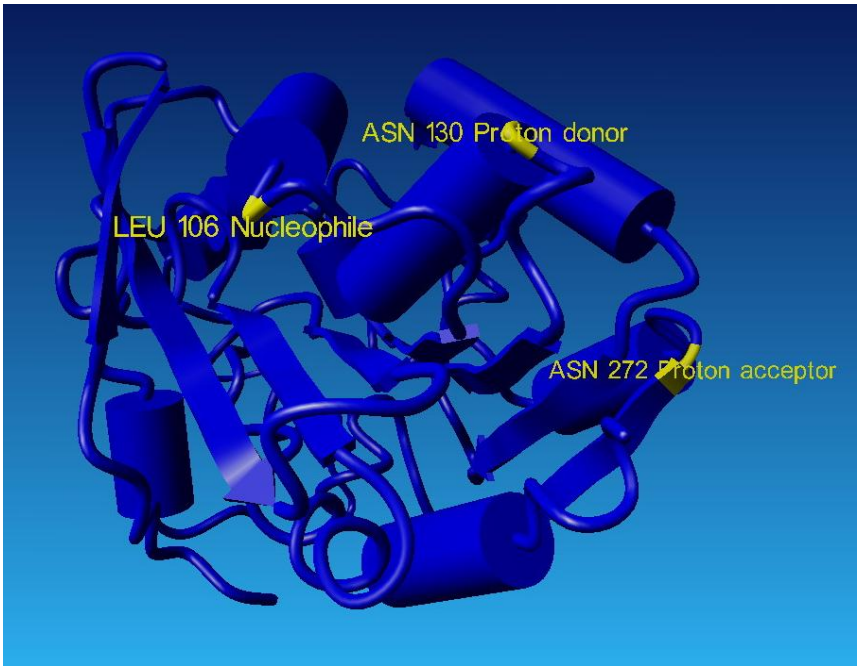

Figure 9. Cartoon model of native DhaA from *Rhodococcus rhodochrous*

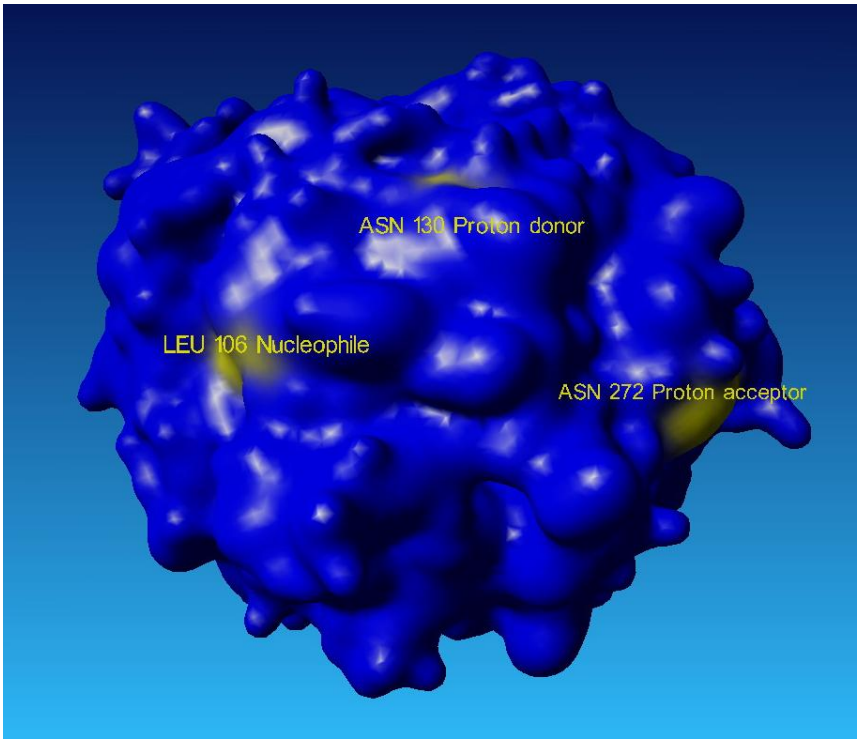

Figure 10. Surface model of native DhaA.

All those models were made using the Yasara program [24].

The cleavage of the carbon-halogen bond in 1-chlorobutane, which is the key step in its catabolism, is catalyzed by this DhaA dehalogenase and results in the formation of n-butanol. This intermediate is subsequently oxidized in two steps to n-butyric acid (Fig. 11), which can serve as a growth substrate for many bacteria.

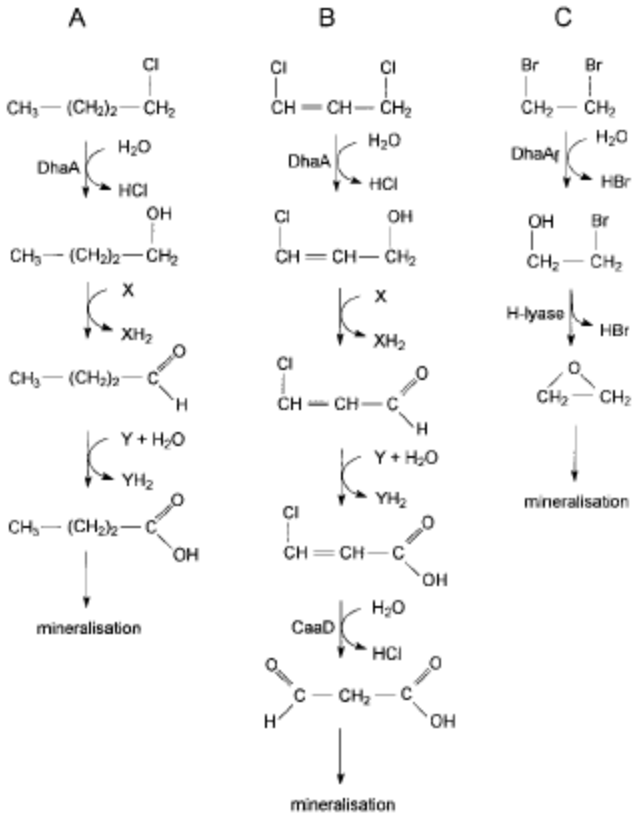

Figure 11. Catabolic pathways for haloaliphatics. [21]

A. 1-Chlorobutane in *Rhodococcus rhodochrous* NCIMB13064.
B. 1,3-Dichloropropene in *Pseudomonas pavonaceae* 170.
C. 1,2-Dibromoethane in *Mycobacterium sp.* strain GP1.

Abbreviations:

DhaA and $DhaA_f$, haloalkane dehalogenases;
H-lyase: halohydrin halogen-halide lyase;
CaaD: 3-chloroacrylic acid dehalogenase;
X: alcohol dehydrogenase cofactor;
Y: aldehyde dehydrogenase cofactor.

The short-chain 1-chloroalkanes (C_3 - C_8) are metabolized by the initial action of a hydrolytic dehalogenase to produce the corresponding alcohol. However, growth on long-chain 1-chloroalkanes (C_{12} - C_{18}) results only in a low level of dehalogenase activity. The attack on the long-chain 1-chloroalkanes is initiated by oxygenase action at the non-halogenated end to produce ω-chlorofatty acids. These are then degraded by β-oxidation.

8.2. The DhaA 12 mutant

The DhaA12 mutant of DhaA from *Rhodococcus rhodochrous* NCIMB13064 was chosen for my crystallization experiments. The aim of creating the DhaA12 enzyme was to enhance the enantioselectivity of DhaA to the extent of the DbjA enantioselectivity, because DbjA shows significant enantioselectivity with selected substrates, e.g. 2-bromopentane in contrast to DhaA that shows no enantioselectivity [28]. Enantiomers are substances with an identical chemical composition, but different space order and it is very problematic to separate them because of their identical physical and chemical properties. The DhaA12 was the most active cumulative mutant towards 1,3-dibromopropane (DBP), which exhibits perfect enantioselectivity towards ethyl 2-bromopropionate and increased enantioselectivity towards 2-bromohexane and 2-bromoheptane in comparison to DhaA, enantioselectivity towards 2-bromopentane was not changed.

DhaA12 differs from the native DhaA by an extra sequence from DbjA ($_{139}$HHTEVAEEQDH$_{149}$) and its entrance and active site changed bias DbjA. It means that DhaA12 contains besides extra sequence (and 305*ins*E (*due to cloning*)) additional point substitutions: W141F, P142A, F144A, V245A, A172V, C176G, G171R and K175G. The protein with molecular weight of 35352.13 Da consists of 311 amino acids. Theoretical pI is 5.18.

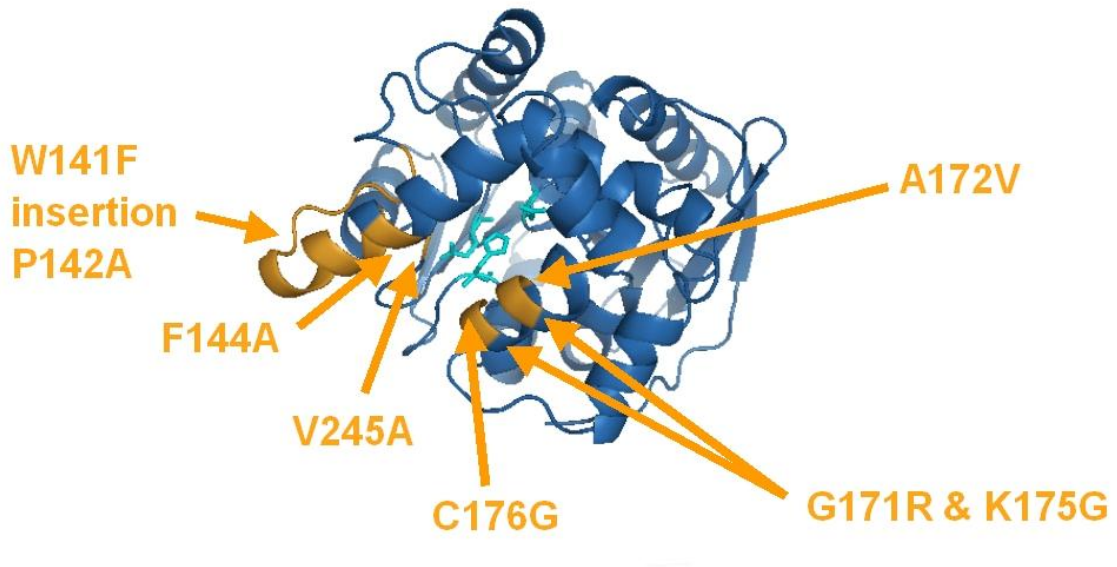

Figure 12. Suggested structure of DhaA12 dehalogenase with sites of mutations shown on the picture.

II. RESULTS AND DISCUSSION

1. Crystallization of the DhaA12 protein

There are many crystallization techniques (e.g. [1, 2, 3, 4, 5, 7]), but for our crystallization experiments only one of them, the sitting drop variant of vapor diffusion technique was used.

Commercial screening kits marked as Crystal Screen™ (HR2-110) and Crystal Screen Cryo™ (HR2-112) from Hampton Research (Hampton Research, CA, USA) were used for initial screening of crystallization condition, but with no success in fact. Then we tried to find similar proteins by sequence similarities using FASTA protein alignment tool (http://www.ebi.ac.uk/fasta33/). The most similar ones were DhaA dehalogenases from various organisms with PDB ID 1BN6 and 1BN7 and the LinB dehalogenase [19], so we tried to refine the conditions used for crystallization of those proteins. This approach seemed to be later successful.

Crystallization procedure describing step by step:

1. 1000μl of precipitant solution was placed into a reservoir.
2. A drop of 1μl protein mixture was pipetted directly into the little depression of crystallization plate.
3. 1μl of a reservoir solution was added to a protein mixture droplet.
4. Steps 1-3 were repeated until all experiments were performed.
5. A crystallization plate was covered with two stripes of the crystal clear tape to maintain the closed system in every single reservoir.

1.1. Crystallization conditions

The crystallization conditions found for crystallization of a native DhaA dehalogenases in their PDB files are as follow:

1 μl protein solution consists of:

- 5 mg/ml protein
- 150 mM ammonium sulfate
- 10 mM TrisSO_4 pH 7.5
- 1 mM EDTA

Reservoir solution consists of:

- 20-25% polyethylene glycol (PEG) 1.500
- 0.2M natrium acetate
- 0.1M MES pH 5.5

Temperature during crystallization experiments was 8°C.

After several crystallization experiments the protein solution was found to be ideal and was used for consequent experiments in following composition:

- 5 mg/ml protein
- 150 mM ammonium sulfate
- 25 mM TrisHCl
- 1 mM EDTA

In each experiment 1 μl of protein solution and 1 μl of reservoir solution were mixed in the little depression of crystallization plate. Other ratios protein:precipitant (1:2 and 2:1) were tested as well, but the best results were achieved with the 1:1 ratio.

Next aim of crystallization experiments was aimed to find appropriate reservoir solution. As a precipitant a PEG of various molecular weights were tested. First we tried 20 - 25% solutions of little molecular weight PEGs such as PEG 550, 1000, 1500 or 2000, according to the initial crystallization conditions and we found that PEG 6000 is the best precipitant. We also tried various concentrations of MPD but with no success. At the end several salts as natrium acetate, ammonium sulfate, magnesium sulfate, calcium chloride and others in concentration from 0.1M to 0.4M were tested. The best results were showed with using of 0.1M natrium acetate.

As a buffer using the MES (2-morpholinoethanesulfonic acid) with pH 6.5 and pH 5.5 resulted to bad crystal composition, similarly a HEPES-Na buffer with pH 7.5 that yielded to no crystals. The best results were obtained using 0.1M Tris HCl buffer with pH 8.5 to 9. The best composition of our reservoir solution was following:

- 20% PEG 6000
- 0.1M natrium acetate
- 0.2M ammonium sulfate
- 0.1M TrisHCl pH 8.5 – 9

The temperature maintained during this experiment was 8°C.

1.2. Results

Crystallization experiments performed in Linbro plates (Hampton Research, CA, USA) yielded thick needles crystals of various lengths (Fig. 13, 14). Crystals with dimension of about 0.5 to 1×0.1×0.1 [mm] appeared after 3 days at temperature 8°C.

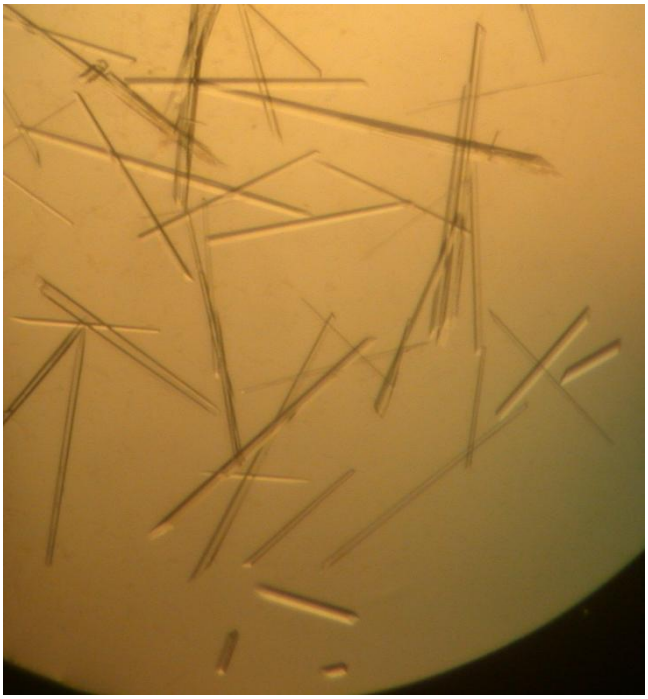

Figure 13. Crystals of DhaA grown from buffer TrisHCl pH 9.0

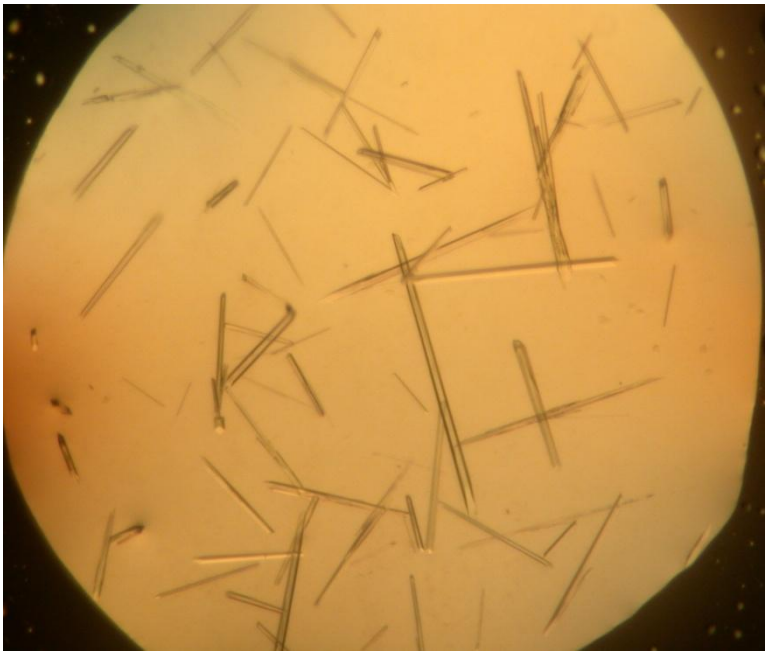

Figure 14. Crystals of DhaA grown from buffer TrisHCl pH 8.5

2. Testing of crystals

Crystals were tested at home diffractometer at the Institute of Molecular Genetics AS CR in Prague. Preliminary diffraction measurements clearly showed protein character of crystals.

3. The diffraction experiment

To carry on diffraction experiment the crystal was mounted into a cryoloop (Hampton Research, CA, USA) and then rapidly frozen by immersing into liquid nitrogen. The loop was then fixed into the goniometer head, where constant nitrogen stream from Nitrogen Gas Low Temperature Cryostat (by Oxford Cryosystems) was cooling down the crystal. The crystal was set into the middle of the X-rays path, so that the diffractions were most accurate. The rotating anode Nonius FR591 (by Bruker-Nonius) was used as X-ray source. The wavelength of the X-ray was 1.54179 Å. We set the measuring time at one angle to 15 minutes. We measured 90° that means 90 turns of the goniometer head of 1°. The image plate Mar345 (MarResearch, [20]) was used as X-ray detector. After every turn the picture on the image plate was scanned with a laser, transferred into the computer and saved in special file for further processing.

4. Data processing and molecular replacement

The processing of digitalized images follows after data collection. This was done using the MOSFLM program. Subsequent procedure is described below.

First we ran autoindexing, where the unit-cell parameters were determined and the possible space group suggested. The program gave us several suggestions with their penalties (the lower penalty, the higher concordance of a space group with the measured data), space group of our analogue 1BN6 was P 2_1 2_1 2, but we used the P 2_1 2_1 2_1 space group (orthorhombic), which had a lower penalty. The unit cell size was a = 52.739Å, b = 68.900Å and c = 84.695Å. Then we let the program refine various parameters like crystal-detector distance, direct beam position, exact position of the crystal and estimate mosaicity of the crystal is estimated.

The last step was the integration, where the output files were built. The reflection spots on the integrated images were converted into intensities and recorded in the MTZ file (a map file). Autoindexing or refining cell parameters were written into a file containing the refined crystal orientation matrices.

5. Data scaling

Next we scaled the data and converted intensities into structure factors.
The scaling was done using a program called "SCALA". This program scales together multiple observations of reflections and merges multiple observations into an average intensity. We let 5% of the data aside from the scaling, because they served as the "air data" – the blank for determining of the R free factor. Then we ran SCALA and the output was provided to the program named "TRUNCATE". This program converts reflections' intensities into structure factors.

6. Phase problem solving

The phase problem was solved using molecular replacement (MR) as sufficiently analogous proteins are available. We used the "MOLREP" program. The input was the MTZ file, which was output by "TRUNCATE" and the PDB file of the protein analogue. Our analogue was 1BN6, which showed 94.737% similarity with our DhaA12 protein, according to the FASTA search in the UniProt database [25]. The output was then a PDB file of our protein.

The last part of this stage was the rigid body refinement. We ran "REFMAC5", set the "rigid body refinement" and in 5 cycles we obtained the resolution of ca 84Å – 3Å. Then electron density maps were computed from the resulting MTZ file that was converted to the structure factor map (HKL file in CIF format) and was used as a map input for "XFIT" (Fig. 15).

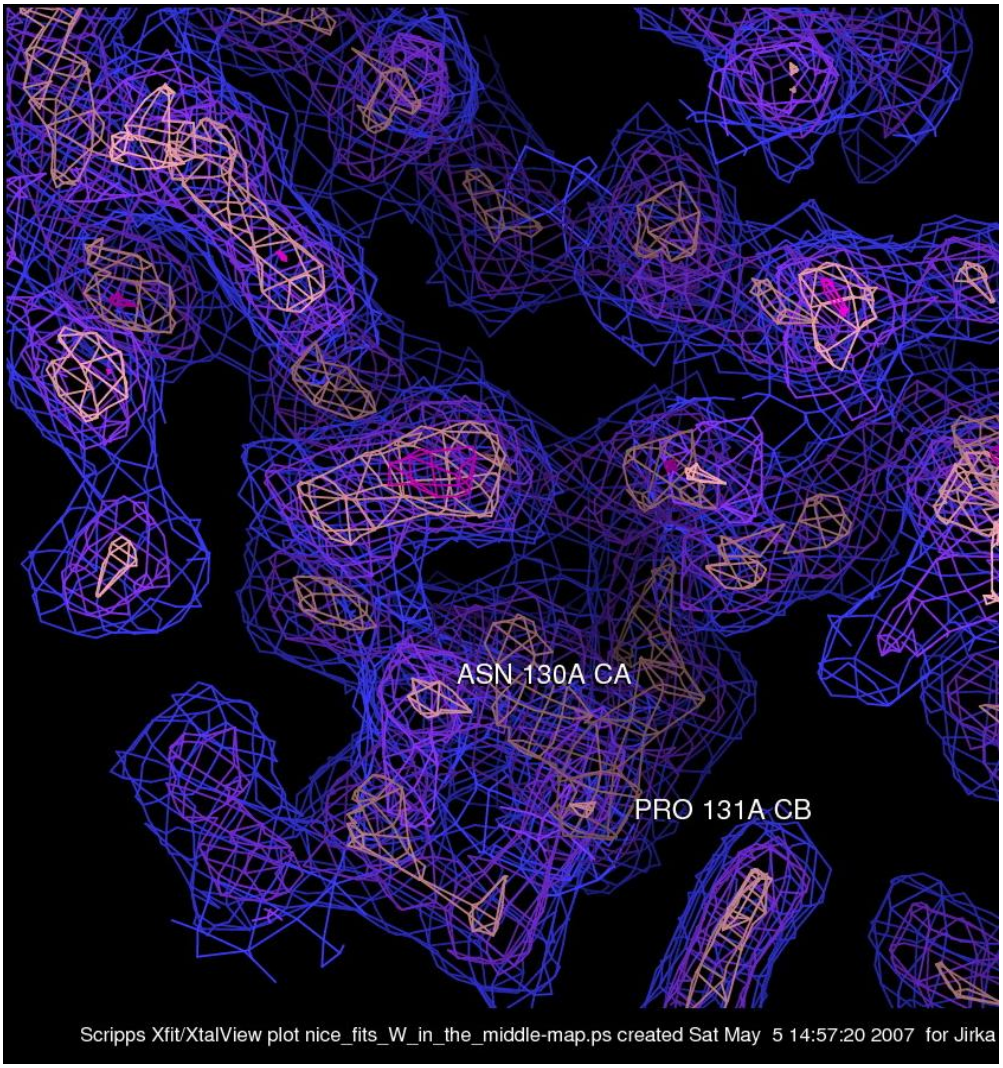

Figure 15. Image from XFIT, where good electron density maps are displayed.

7. Amino acids fitting and structure refinement

Amino acids fitting and structure refinement are the last and most interactive phase. We used programs – "XFIT" and "REFMAC5" in pair. In XFIT we fit the amino acids into the model. As the 1BN6 protein showed high similarity to our DhaA12, at the beginning we had an almost complete model with amino acid residues fitted, but we had to build and fit the filled in side chain and also fix some point mutations. After fitting the amino acids we ran the "REFMAC5" to correct some parts of electron maps or the model (e.g. the disulphidic bonds). We can also set the restrained refinement, eventually water molecules filling in this program. Then we created maps in "XtalView" format and let the program run in 5 cycles in the highest resolution possible. With the lower resolution electron density maps become more concrete. When refinement was finished, we returned to "XFIT" again and we tried to improve our

model. It is important because what is not in the electron densities, it should not be in the model and in adverse. When we finished the editing in "XFIT", we returned to "REFMAC5" and ran refinement again. The aim of this step is to have a model, which is the most similar to the real structure of the molecule (Fig. 16). The R factors serve as criteria of right.

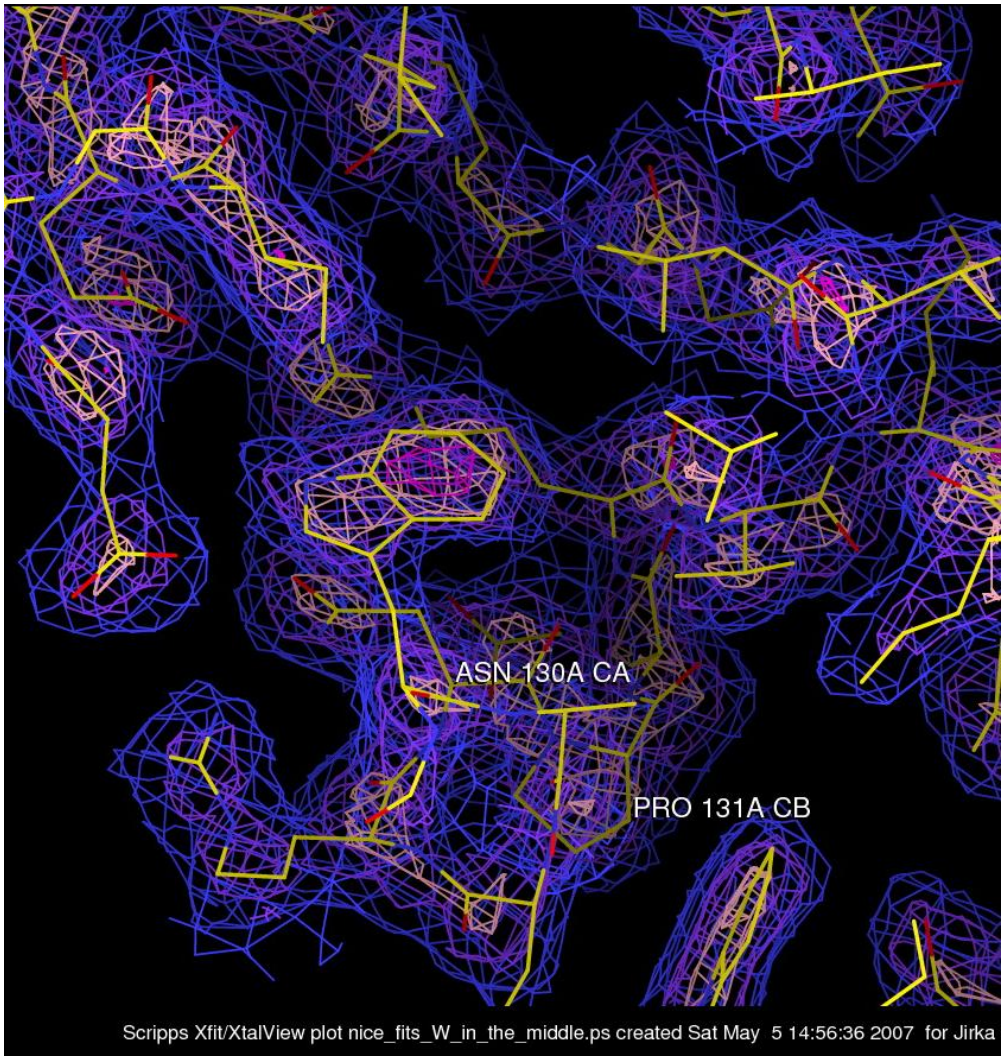

Figure 16. The same image like in Fig. 15 from XFIT, where electron densities and passing amino acids are displayed.

The last phase of refinement was to check the model and find there the last errors. We checked the thermal factors directly in the PDB file. There cannot be very big difference in the same molecule because the crystals are mostly diffracted in the frozen state, so the values over 50 can be immediately deleted.

If we have completed the structure refinement and cannot see any mistakes or errors, we can test our structure using the online validation services. We used the following ones.

- **Procheck:** This software is an integral part of CCP4 Suite [6], version 6.0.2. According to [16], the current checks performed by PROCHECK on given protein structure are as follows: Covalent geometry, planarity, dihedral angles, chirality, non-bonded interactions, main-chain hydrogen bonds, disulphide bonds, stereochemical parameters, parameter comparisons, residue-by-residue analysis.

8. Final Structure

Using procedures described in previous paragraphs, protein structure was finally solved. The main differences between native DhaA dehalogenase and its mutant DhaA12 are showed on Fig. 17.

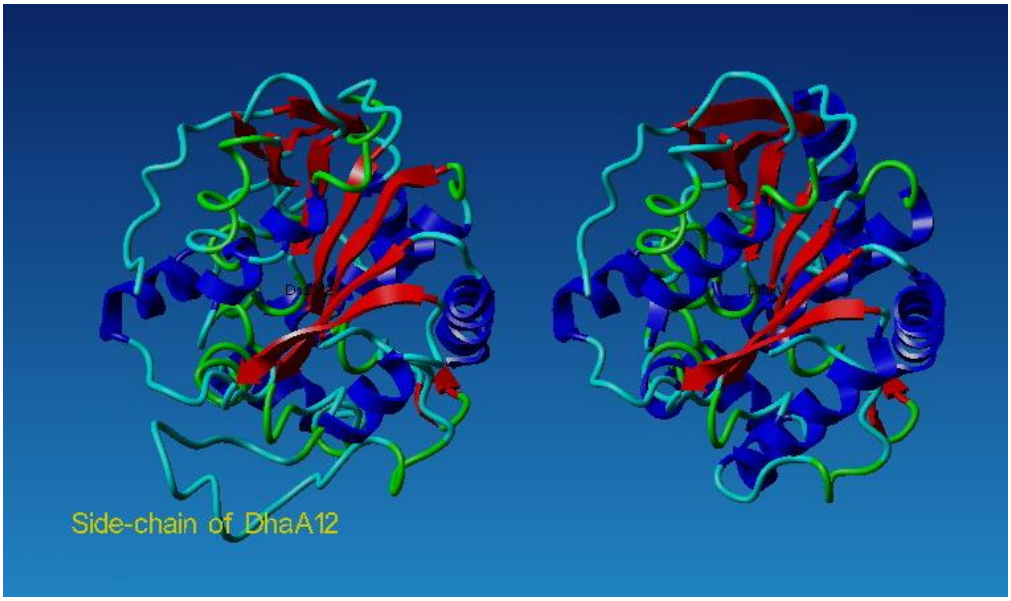

Figure 17. Comparison of DhaA12 (left) and native DhaA (right). The inserted side-chain is marked.

Fig. 18 shows the DhaA12 protein with the active sites, which are supposed to be analogous to the 1BN6 native DhaA. The side-chain loop is situated in the bottom part of the picture.

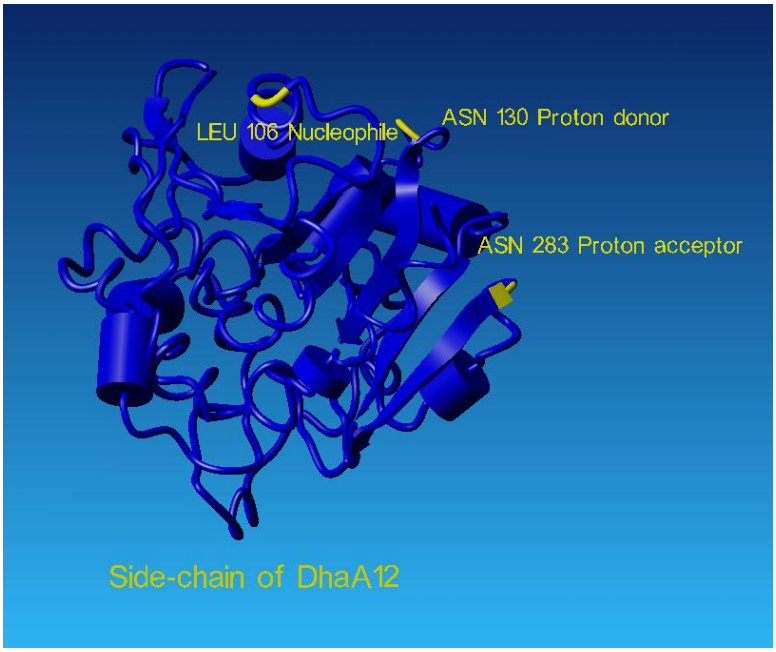

Figure 18. DhaA12 cartoon model with active sites highlighted.

According to Procheck the structure is right. The final R-factor is 0.20, which is acceptable for a right structure and the R_{free} factor is 0.38 that is still quite high. According to various structures checking software the biggest problem seems to be the side-chain. This also suggests a view of the structure in the XFIT program. The whole structure is clear, the electron densities show clearly positions of amino acids, but the problem is that at the place where the side-chain is inserted only few electron densities are present to suggest the direction where the chain should be situated. These scarce densities are located in the chain from Val157 to Asp162 shown on Fig. 19 and then from Ala166 to Arg168, as Fig. 20 shows.

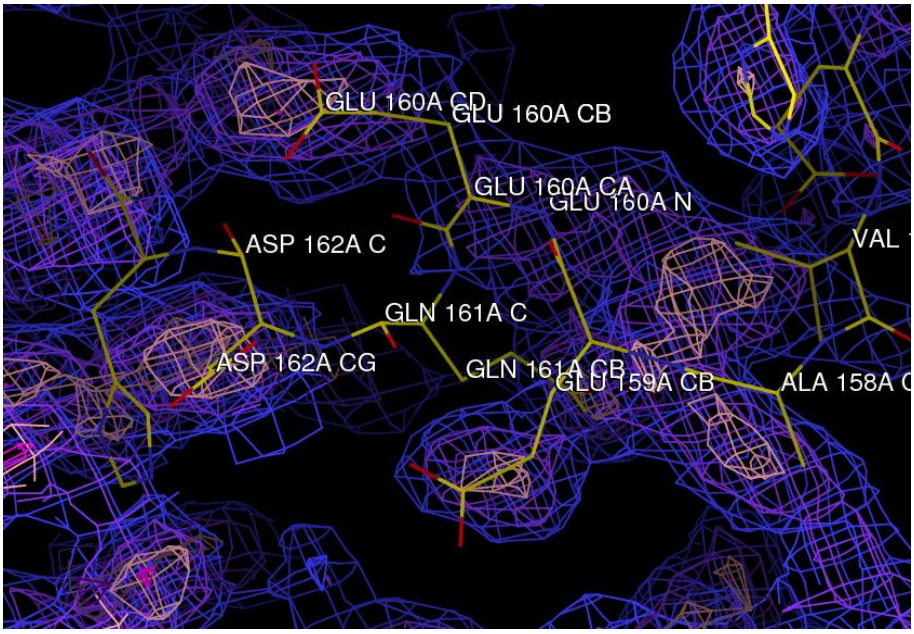

Figure 19. Missing electron densities around Glu159, 160, Gln161 and Asp162

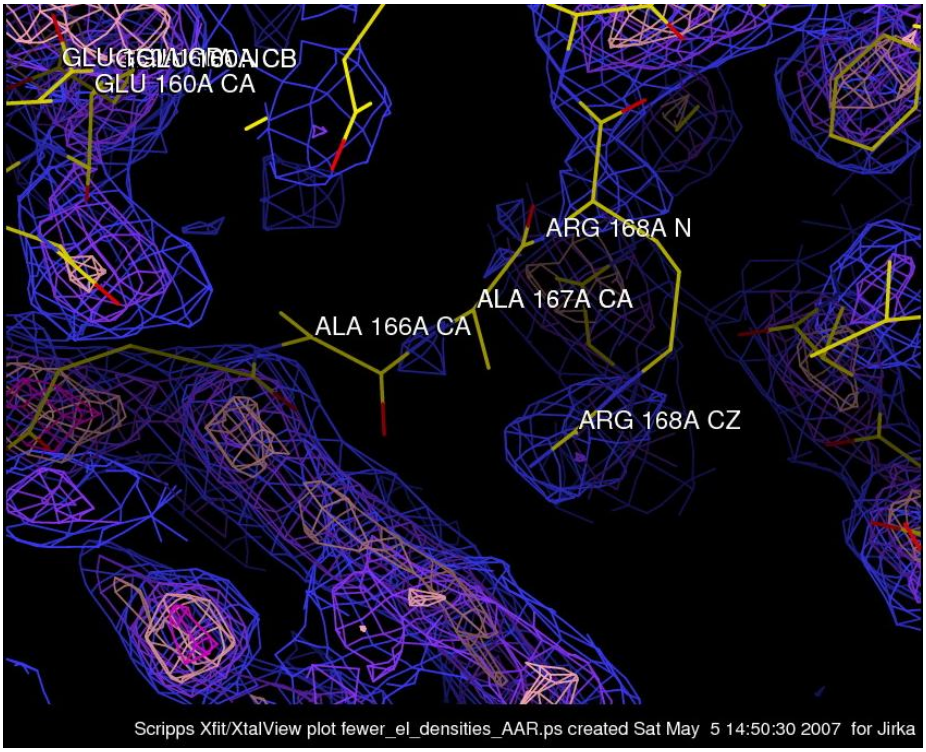

Figure 20. Missing electron densities around Ala166, Ala167 and Arg168

After many attempts the best possible way how to determine the direction was found according to the high electron densities in the middle of the chain. The missing electron densities at the entry and exit of the side-chain from the main molecule made it impossible to determine this exactly. This is probably the reason why the R_{free} factor is quite high, because it most penalizes the parts of the model, which are not present in the electron density map.

There are two reasons why this happened and these cohere with each other. The resolution of measured data was quite low – 3.00Å. Then the inserted side-chain is probably very flexible, so in the crystal some parts of chain are ordered in the other way in each protein molecule and for this reason only a few good reflections of this site were recorded and converted to electron densities. This may also be the reason getting of low resolution. Several crystals were tested before measurement and all of them showed lower resolution, what is probably problem of this free loop.

III. CONCLUSION

All objectives of this work were fulfilled. The DhaA12 mutant of haloalkane dehalogenase was successfully crystallized, test have proven the crystals to be a protein and its structure was solved to the best possible rate. The whole structure is reliable, the only uncertainty lies at the beginning and the end of the inserted side-chain, where it was impossible to fit altogether 7 amino acids correctly, the inner part of the side-chain is clear. Also 7 out of 316 amino acids are possibly misplaced, the rest of the structure stays correct. Fig. 20 shows the ribbon model of DhaA12 mutant haloalkane dehalogenase.

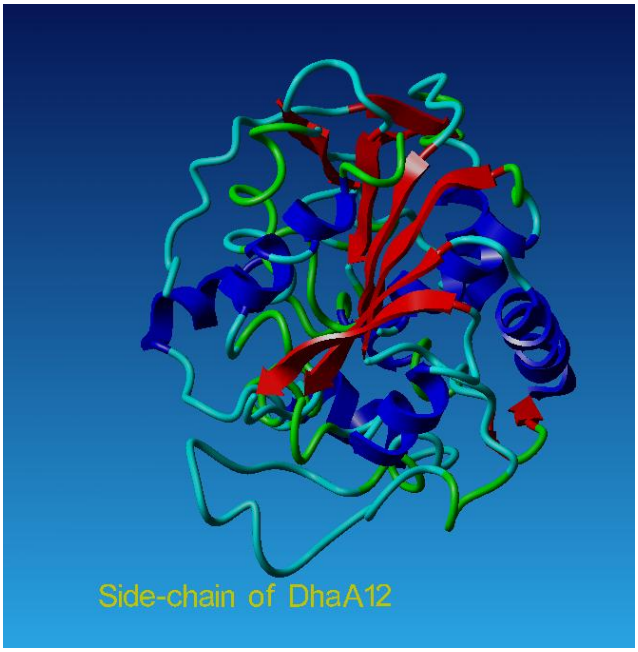

Figure 20. The DhaA12 ribbon model.

IV. REFERENCES

[1] McPherson: Crystallization of Biological Macromolecules. New York 1999. CSHL Press.

[2] Ducruix and R.Giege: Crystallization of Nucleic Acids and Proteins. New York 1999. Oxford University Press, Inc.

[3] T. M. Bergfors: Protein Crystallization: Techniques, Strategies and Tips. La Jolla 1999. IUL Biotechnology Series.

[4] D. M. Bolag et al.: Protein Methods. New York 1996. Wiley-Liss, Inc.

[5] Garcia-Ruiz, J.M., Gonzales-Ramirez, L.A., Gavira, J.A. and Otalora, F.: *Acta Cryst*, 2002, **D58**, 1638-1642.

[6] COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4. 1994. "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. D50, 760-763

[7] J. Drenth: Principles of Protein X-Ray Crystallography, Second Edition. New York 1999. Springer-Verlag

[8] *http://www-structure.llnl.gov*

[9] *http://www.eserc.stonybrook.edu/ProjectJava/Bragg/index.html*

[10] *http://www.mrc-lmb.cam.ac.uk/harry/mosflm/mosflm_user_guide.htm*

[11] *http://xray.imsb.au.dk/~deb/Thesis/node77.html*

[12] *http://www.ccp4.ac.uk/dist/html/refmac5.html*

[13] CCP4 v4.2.2 Program Documentation. http://www.ccp4.ac.uk/dist/html/INDEX.html

[14] XTalView. *http://www.sdsc.edu/CCMS/Packages/XTALVIEW/xtalview.html*

[15] *http://bass.bio.uci.edu/~hudel/mbb254/lecture2/lecture2_1.html*

[16] *http://biotech.ebi.ac.uk:8400/*

[17] *http://kinemage.biochem.duke.edu/molprobity/main.php?use_king=1*

[18] *http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phases.html*

[19] Smatanová, I., Nagata, Y., Svensson, L. A., Takagi, M., Marek, J.: *Acta Cryst.* (1999). D**55**, 1231-1233

[20] *http://www.marresearch.com/ip.htm*

[21] Gerrit J. Poelarends, Leonid A. Kulakov, Michael J. Larkin, Johan E. T. Van Hylckama Vlieg, and Dick B. Janssen: Journal of Bacteriology, Apr. 2000, p. 2191 – 2199

[22] Anna N. Kulakova, Michael J. Larkin and Leonid A. Kulakov: Microbiology (1997), 143, 109 – 115

[23] *http://www.expasy.org/uniprot/P0A3G2*

[24] Krieger E, Koraimann G, Vriend G (2002) Proteins 47,393-402

[25] *http://www.ebi.uniprot.org/index.shtml*

[26] Andrew G. W. Leslie: Acta Cryst. (2006). D62, 48–57

[27] Dick B Janssen, Jantien E Oppentocht and Gerrit J Poelarends: Current Opinion in Biotechnology 2001, 12:254–258

[28] Sato, Y., Monincova, M., Chaloupkova, R., Prokop, Z., Ohtsubo, Y., Minamisawa, K., Tsuda, M., Damborsky, J., Nagata, Y., 2005: Applied and Environmental Microbiology 71: 4372-4379.

[29] Hiroaki Adachi, Kazufumi Takano, Masaaki Morikawa, Shigenori Kanaya, Masashi Yoshimura, Yusuke Moria and Takatomo Sasaki: Acta Cryst. (2003). D59, 194±196

[30] Philip Evans: Acta Cryst. (2006). D62, 72–82

[31] Martin A. Walsh, Gwyndaf Evans, Ruslan Sanishvili, Irene Dementieva and Andrzej Joachimiaka: Acta Cryst. (1999). D55, 1726±1732

[32] *http://www.bchs.uh.edu/struct_bio/MIR_course/*

[33] *http://www.cryst.bbk.ac.uk/pps97/assignments/projects/kozak/PHASE.HTM*

[34] *http://en.wikipedia.org/wiki/Phase_problem*

[35] Qun Shen, Daniel Pringle, Marian Szebenyi, and Jun Wang: Review of Scientific Instruments, Volume 73, Number 3, March 2002

[36] Emmer, J., 2004. Crystallization and Crystallographic Studies of Model Protein, Bc. Thesis, in English. – 39. p., Faculty of Biological Sciences, University of South Bohemia, Czech Republic