# Simulation studies of Coherent X-Ray diffraction Single molecule imaging on XFEL

### Wanaruk Chaimayo

Mahidol University, Thailand

**Prof. Ivan Vartanyants** 

Dr. Oleksandr Yefanov

Supervisors

HASYLAB Group, DESY Summer Student Program 2008

#### ACKNOWLEDGEMENT

First of all, I wish to thank all the people who made completion of project possible. My deepest appreciation goes to my supervisor Prof. Ivan Vartanyants and Dr. Oleksandr Yefanov for their valuable advice and their powerful encouragement throughout this research. I also would like to express my gratitude to A. Mancuso and J. Gulden for their useful comments and excellent suggestions.

I wish to sincerely express my appreciation to Prof. Dr. Jochim Meyer, Andrea Schrader and people who are behind the summer student program are also greatly acknowledged for this invaluable experience. Furthermore, it is my great honour to be one of the Thailand representatives which passed the national selection of NSTDA and NSRC under the patronage of HRH Princess Maha Chakri Sirindhorn.

Finally, we would like to use this opportunity to express my grateful thanks to our family, for their support and understanding.

## CONTENT

Acknowledgement	Page
Abstract	4
1. Introduction	5
2. Experimental Procedures	6
3. Results and Discussions	13
3.1 Simulating Coherent X-Ray Diffraction Imaging	
of TBSV by MOLTRANS	13
3.2 Reconstructing Real Space Image of TBSV from	
diffraction pattern by RECON	14
3.3 Add XFEL photon flux in diffraction image for	
more realistic situation	23
3.4 Reconstructing Real Space Image of TBSV from	
modified diffraction pattern by RECON	23
4. Conclusions	27
5. Further Study	27

# Simulation studies of Coherent X-Ray diffraction Single molecule imaging on XFEL

Wanaruk Chaimayo<sup>*a*</sup>, Oleksandr Yefanov<sup>*b*</sup>, Ivan Vartanyants<sup>*b*</sup>

<sup>a</sup> Summer student; Department of Physics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

<sup>b</sup> Supervisor; HASYLAB, DESY, Notkestrasse 85, D-22607 Hamburg, Germany

#### Abstract

X-ray Free Electron Laser (XFEL) was proposed to use as the x-ray source in coherent x-ray diffraction imaging (CXDI) for single molecule structure determination<sup>1</sup>. This study was performed to find the probability to get the reasonable outcome from CDXI under XFEL properties, such as number of photon per bunch at the particular wavelength. This work simulated Fronhofer x-ray diffraction of a certain virus molecule, and then reconstructed the diffraction pattern to the real space of the virus by phase retrieval method. As a result, this simulation study demonstrated that there is the possibility to use XFEL as the X-ray source for coherent x-ray diffraction imaging (CXDI) to determine Single Molecule structure at 1 angstrom x-ray wavelength, 1 angstrom resolution and 10<sup>12</sup> photons per bunch corresponded to XFEL performance goals.

<sup>&</sup>lt;sup>1</sup> A.Schropp, I. Vartanyants, IPAP Conf. Series 7 pp.383-385

# Simulation studies of Coherent X-Ray diffraction Single molecule Imaging on XFEL

#### **INRODUCTION**

The European X-Ray Free Electron Laser (XFEL) facilities which will be built in 2008 near DESY, Hamburg Germany, to be operational in 2013, is expected to have a strong impact in a wide range of scientific domains synchrotron radiation facilities because it will generate extremely brilliant and ultra short pulse (~100 fs) of spatially coherent X-ray in a wave length regime from 0.1 nm to 6.4nm. The performance goal for FEL radiation in term of number of photon per bunch is in order of  $10^{12}$  for 1 angstrom wave length. <sup>2</sup>

Coherent X-Ray diffraction imaging<sup>3</sup> of Single Molecule, part of molecular structure study, needs the outstanding properties of XFEL which are femtosecond pulse and ultrahigh brilliance. This special x-ray pulse is not going to destroy the single molecule before finishing the measurement and strong enough to make diffraction pattern on single molecule sample.

Consequently, these performances of XFEL can make the single molecule imaging possible. During the construction of XFEL, we made a set of simulation to determine the possible result of the experiment of Coherent X-Ray diffraction Single molecule imaging on XFEL. The simulation use Tomato Bushy Stunt Virus as the single molecule for X-ray imaging.

<sup>&</sup>lt;sup>2</sup> TESLA XFEL – An Executive Summary

<sup>&</sup>lt;sup>3</sup> A.Schopp, I. Vartanyants, et al, 8<sup>th</sup> Int. Conf. X-Ray Microscopy, Conf. Series 7 pp.383-385

#### **EXPERIMENTAL PROCEDURE**

We have conducted the project by the following steps.

2.1 Simulating Coherent X-Ray Diffraction Imaging of TBSV by MOLTRANS

2.2 Reconstructing Real Space Image of TBSV from diffraction pattern by RECON

2.3 Add XFEL photon flux in diffraction image for more realistic situation

2.4 Reconstructing Real Space Image of TBSV from modified diffraction pattern by RECON

#### 2.1 Simulating Coherent X-Ray Diffraction Imaging of TBSV by MOLTRANS

Tomato Bushy Stunt Virus (TBSV), well known virus in Tomato, was studied by Harrison S.C.<sup>4</sup>. TBSV data is provided as Protein Data Bank (PDB) file. We can download this particular file in RCSB Protein Data Bank website.



(From http://www.rcsb.org)

Fig. 1. Tomato Bushy Stunt Virus has a spherical shape and 30 nm diameter<sup>5</sup>

We use this certain virus for the simulation. First, we need to know how the diffraction image of this molecule is. As a result, for the first step, we performed simulation under the far field condition and unlimited photon flux to detector (ideal case) by using MOLTRANS.

 <sup>&</sup>lt;sup>4</sup> J.Mol. Biol. 177:701-713 (1984)
 <sup>5</sup> T. Ohki J Gen Plant Pathol ,**VOL.**71;**NO.**1;**PAGE.**74-79(2005)

MOLTRANS is the FORTRAN language program which has been developed by E. Weckert<sup>6</sup>. This program can be used for simulating the scattering pattern of any molecule which is represented in PDB file.

We use MOLTRANS (revision 1.9) via UNIX to simulate scattering image in hascl0g computer cluster, DESY.

#### List of Common used Parameter in MOLTRANS

(The used input parameters is provides in [] \*for TVSB case only)

Enter file name of coordinate (pdb) [filename.pdb] Found BIOMT symmetry 0

Apply virus symmetries, not needed for BIOMT (y/n) [n] Enter number of Molecular Shell (e.g. 2 3 5) [1 1 1]

Calculation for detector (d) or in rez. space (r) (d/r) [d] Read molecule translation from file (y/n) [n] Change periodicity for calculations (y/n) [n]

Enter Wavelength in A [depends] Enter sample detector distance in mm [from calculation] Enter Detector Size in mm [depends] Maximum resolution is (given by the program) Enter number of detector pixels (nx,ny) [depends]

Apply temperature factor? 0=none, 1=yes, -1=bo [0]

Rot(z)  $\Psi$  rotaion: start, end, number [0 0 1] Rot(x)  $\Theta$  rotaion: start, end, number [0 0 1] Rot(y)  $\zeta$  rotaion: start, end, number [0 0 1]

<sup>&</sup>lt;sup>6</sup> HASYLAB, DESY, Hamburg, Germany

Enter sig. cm of beaml in A(x,y) [0 0 0 0] Enter output file name [fileoutname]

As for the output of MOLTRANS, it is binary file. To make this more perceptual image, we have to convert it to ascii file and, finally, convert to tiff file version by additional program which are describe below.



Fig. 2. Describe the process to get the simulated diffraction pattern

For converting to tiff file format, we use dir\_to\_ascii to change binary file from MOLTRANS and ascii\_to\_tiff for bring about tiff picture from ascii file input.

#### List of Common used Parameter in Dir\_to\_ascii

```
Enter input fileman [fileoutname]
.
.
.
Enter (next) input file name (none for continue) [] //just enter, no more file
File outputname [fileoutnameasc]
Log or linear (log/lin) [lin]
Enter bining Factor in x and y [1 1]
Enter z-maximum for plot [1e38] [] //just enter
Enter z-minimum for plot [0] [] //just enter
```

#### List of Common used Parameter in Ascii\_to\_tiff

Type in the terminal

>> ./ascii\_to\_tiff fileout**asc** fileout**tiff** 

Then we can have a quick look with **Kuickshow** in UNIX by typing the following command in the terminal

>> Kuickshow fileout**tiff** 

#### **Calculation for MOLTRANS input parameters**

We have to calculate some parameters which are according to physics of X-ray scattering principle by the following formula.

$$R = \frac{Z \bullet \lambda}{D}$$

When

Required Resolution (R) Distance between detector and sample (Z) Wavelength of X ray ( $\lambda$ ) Detector Size (D)

#### 2.2 Reconstructing Real Space Image of TBSV from diffraction pattern by RECON

RECON program was created by O.Yefanov<sup>7</sup>. This program is practically used to invert scattering images of any samples to their real space images. Powerful phase retrieval algorithm was used as the key method of the program.

To obtained the good quality reconstruction image, we have to be seriously concerned about some parameters which were used in this program, e.g. support size, number of iteration, positive, to name but a few.



Fig. 3. Overview of the simulation steps

To reconstruct any scattering image to real space image, we have to modify the scattering picture by ImageJ, scientific image editor program, in the following steps.

- 1. Get the particular scattering image which is in \*.tif file format
- 2. Change type of image from 8 bit to 32 bit
- Switch off the box in Edit->Options->Input->Output-> Save TIFF and Raw in Intel Byte Order
- Save this modified image as \*.RAW file Nevertheless, RECON program always need *libfftw3-3.dll* file and *conf* file. The *conf*

<sup>&</sup>lt;sup>7</sup> HASYLAB, DESY, Hamburg Germany

file is the RECON's configuration text file which is vital for reconstruction process. Details information of the configuration file is described below.

After we open a *conf* file with a text editor, we will see (The explanations are provided after //)

SizRec: 500 500	// number of pixels (x,y) of detector we used in MOLTRANS
SizSup: 70 70	// the reconstruction image will be in $70x70$ (pixel x pixel) frame
BigEndian: 1	
files: 9tiff_tvsb_lin.raw	// the input *.RAW file after modified by ImageJ
Positiv: 1	
Process: 1	// amount of real image we need
CropImage: 0	// crop image
er 30 1 0.8 0	// er and hio are the iteration process for phase retrieval method
hio 100 1 0.8 0	// hio 100 means 100 iteration
er 30 1 0.8 0	// er 30 means 30 iteration
hio 100 1 0.8 0	
er 30 1 0.8 0	
hio 100 1 0.8 0	
er 30 1 0.8 0	
hio 100 1 0.8 0	
er 30 1 0.8 0	// end

For configuring the reconstruction image, we have to put the essential parameter which shown above in the *conf* file and save the file after finish. Then we have to put all file, which are *RECON.exe*, *libfftw3-3.dll*, *fileouttiff.raw* and *conf*, in the same folder. Finally, to run RECON program, click the program and then press space bar<sup>8</sup>, the program will start reconstruction process automatically.

<sup>&</sup>lt;sup>8</sup> depends on available versions

#### 2.3 Add XFEL photon flux in diffraction image

As for the first step, we simulate the ideal case of photon flux. However, to perform more realistic situation, we introduce the absolute photon flux of XFEL to the detector which is  $10^{12}$  photon per bunch for 1 angstrom wave length <sup>9</sup>.

As for this purpose, we use to TOCPS program, created by E.Weckert<sup>10</sup>, to add photon flux number. We simulated  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$ ,  $10^{15}$  and  $10^{16}$  photons per bunch cases at 1 angstrom diffraction pattern.

There are some steps to modified limited photon situation from ideal photon case. We need ascii file from dir\_to\_ascii to be an input. And, output from TOCPS is ascii file as well.

#### List of Common used Parameter in TOCPS

(all processes are performed in UNIX)

type ./tocps in the terminal to get access to the program

enter distance [mm] 50// as we use for simulating particular diffraction patternenter pixel size [mm] 0.0244140625// as we use for simulating particular diffraction patternenter no. of photons/bunch [ph] 10e11// Beam properties of European XFELenter focal spot [um] 0.1// Beam properties of European XFELflux density in ph/mm^2, scale 0.10000E+22 0.18826E-08 ok ? (y/n) [y]maximum intensity for plot 4e9plot only positive quarter (y/n) [y] nEnter input file name 1ascii\_tbsv// ascii file which you get from dir\_to\_asciienter output file name 1ascii tocps tbsv

Next process, we use ascii\_to\_tif as describe in step 1 to get diffraction pattern in \*.tif file format.

# 2.4 Reconstructing Real Space Image of TBSV from modified diffraction pattern by RECON

For reconstructing, we also follow the steps as indicated in step 2.2

<sup>&</sup>lt;sup>9</sup> TESLA XFEL- An Executive Summary

<sup>&</sup>lt;sup>10</sup> HASYLAB, DESY, Hamburg, Germany

#### **RESULTS AND DISSCUSSIONS**

#### 3.1 Simulating Coherent X-Ray Diffraction Imaging of TBSV by MOLTRANS



Fig. 4. Simulated diffraction patterns of TBSV at various parameters



3.2 Reconstructing Real Space Image of TBSV from diffraction pattern by RECON

Fig. 5. The diffraction pattern is shown on the linear scale

At first, we simulated 1.5 angstrom because we have used generally in X-Ray experiment in Synchrotron Source. As for the resolution, 5 angstrom is enough for resolution in this diffraction pattern (500x500 pixel x pixel). The more resolution we use, the more time MOLTRANS have to use to process.

Reconstruction Image with various parameters								
Time	Picture	Filename	Sizeup Pixel(x,y)	Bigendian	Positive	Process	Crop	Itteration
1	Ó	152	80x80	1	1	1	0	4x130
2		144	70x70	1	1	1	0	4x130
3		145	70x70	1	1	1	0	8x130
4	Ô	142	70x70	1	1	1	0	12x130
5		146	70x70	1	1	1	0	16x130
6	<u>(</u> )	15	70x70	1	0	1	0	16x130
				0:Invalid Floating		0 : error at the end		

*Fig. 6.* Reconstruction Images and corresponding parameters

At the first period of project, we have to vary some parameters to know the performances of RECON program. As a result, this table shows the comparison of the reconstruction images which were created under different combination of various parameters.

To improve the quality of Real Space image, we need to change some parameters in reconstruction program which we will describe in 1-4

#### 1. Cropping some area of TBSV Scattering Image.

We have an idea to crop the diffraction image before the reconstruction

process.



Fig. 7. The difference of positive makes the difference of reconstruction image



Fig. 8. The more cropping in the diffraction pattern, the smoother reconstruction images are.

The shapes of the virus in reconstruction images look round. There are some differences, however, in smoothness.

#### 2. Comparison Reconstruction Image between Positivity=1 and Positivity=0

For this case, reconstruction image were obtained from no cropped diffraction patterns. We provided three reconstruction images per one condition to see the reproducibility of the RECON Program.



Fig. 9. Reconstruction image from MOLTRANS output and RECON program

As for comparison, we also simulated TBSV image from MOLTRANS output by getting the data set of amplitude and phase represented in diffraction image. Next, do Inverse Fourier transformation to obtain real space image of TBSV.

From the set of pictures in each condition, the viruses in all reconstructed images are round and look almost the same which shows the stable output of RECON. However, all results are not as good as the expected image from MOLTRANS.

#### 3. Average all reconstruction images

As we see in the last section, the three reconstruction images in a condition look similar, but are not the same. Consequently, we need to average the reconstruction image to be the representation for each condition. First, we again simulated scattering pattern by MOLTRANS with 0.1 Angstrom and 1 Angstrom Resolution.



Fig. 10. the scattering picture is show on the linear scale



*Fig. 11.* Average all reconstruction images by Summation of Complex FFT and Manually Superposition

Turning to consider the image average method, in this case, we had five virus reconstructed pictures. Then, we performed Complex Fourier Transformation for each images. Next, add all complex FT images in 32 bit mode. Finally, do Inverse Fourier Transform. As a result, we got the average real space image back.

When we compared the newly averaged picture with TBSV image we get from MOLTRANS, we could see that they are almost the same. All in all, the average method by doing complex FT is the good method for making the representation picture.

#### 4. Modified Support Shape

As per the support shape in reconstruction process, it is the shape which indicates that where the virus should sit or what shape of the TBSV should be.

For the early section of results, from the reconstruction pictures, we have seen that the shape of the virus is about round and diameter about 240 Angstrom, measured and calculated from ImageJ. As a result, we needed round support to let recon perform a better calculation due to knowing the position and shape of the virus. This led us get the better real space results.

# Round Support vs Square Support







<b></b>				
Iteration	10400			
Positive	1			
Support Shape	Square			
Support Size	180x180	(pixel	x	pixel)
DoSqrt	0			



*Fig. 12.* Real space images which were reconstructed from round and square support (Top) Scattering Pattern in linear scale was used for reconstruction in this case (Bottom)

#### 3.3 Add XFEL photon flux in diffraction image for more realistic situation

In this part, we simulated photon statistics by TOCPS in diffraction pattern (as described in 2.3) which are show in the following section.

# 3.4 Reconstructing Real Space Image of TBSV from modified diffraction pattern by RECON

We obtained real space images of virus molecule in each case of number of photons per bunch.

From Fig.13., we could see that after we applied photon statistics from  $10^{16}$  photons per bunch until  $10^{12}$  photons per bunch which is the upper limit performance of XFEL at 1 angstrom wavelength, the less photons per bunch, the worse reconstruction images were.

However, all reconstructed virus pictures are round and showed some details inside which was good compared with TBSV real space image from MOLTRANS.

In addition, the reconstructed picture at the expected performance of XFEL at 1 angstrom,  $10^{12}$  photon per bunch, was also demonstrated virus details significantly compare with TBSV image from MOLTRANS. Nevertheless, the resolution of this reconstruction images was worse than the one from  $10^{15}$ - $10^{16}$  photons per bunch due to the less information in diffraction pattern.

From  $10^{12}$  photons per bunch diffraction pattern, we could see the great black area which showed the intensity, zero. This led RECON program to obtain less result, about 10 times less than the one from  $10^{15}$ - $10^{16}$  photons per bunch, by measuring the width of the diffraction pattern with ImageJ.

Consequently, from Q-space and Real-space relation below

$$\Delta q \propto \frac{1}{\Delta x}$$

If  $\Delta q$  is reduced ten times,  $\Delta x$  will be increased ten time as well, to satisfy the relation. In this case, the resolution was reasonably increased to 10 angstrom (from 2 angstrom) as you could see from the coarseness of virus image.



*Fig. 13.* Diffraction and reconstruction image in  $10^{12}$ - $10^{16}$  photons per bunch cases for 1 angstrom diffraction pattern

### Additional Simulation and Reconstruction at 5 angstrom X-Ray and 5 angstrom Resolution

For better reconstruction image, under XFEL number of photons per bunch condition, we needed more of this number to make brighter diffraction image which led the finer reconstruction picture.

Turning to consider the performance gold of XFEL, we knew that if we chose the longer wave length we would get more number of photons per bunch.

In this case, we chose 5 Angstrom wavelength to get about  $2x10^{13}$  photons per bunch from calculation. Then we also use TOCPS to simulate the photon statistics.

Electron energy range	10-20 GeV
Electron bunch length (rms)	80 fs
Electron bunch charge	l nC
Normalized emittance	1.4 mrad mm
Uncorrelated energy spread (rms)	2.5 MeV
Photon pulse length (FWHM)	100 fs
Photon energy	0.2; 3.0; 12.4 keV
Wavelength	6.4; 0.4; 0.1 nm
Number of photons per bunch $\times 10^{12}$	430; 20; 1.2
Peak brilliance $\times 10^{33}$	0.06; 1.7; 5.4
Peak power	135; 100; 24 GW

Fig. 14. XFEL performance goals<sup>11</sup>

<sup>&</sup>lt;sup>11</sup> J. Felhaus, Physicsa Scripta. Vol. T110, 413-419, 2004

![](_page_25_Figure_0.jpeg)

*Fig.15.* Diffraction and reconstruction image in  $10^{12}$ ,  $2x10^{13}$  photons per bunch and no photon statistics cases for 5 angstrom diffraction pattern

From the result, we could see some details in the reconstruction image under  $2x10^{13}$  photons/bunch condition compared with those with no photon statistics ( ignore limited number of photon case) and TBSV image from MOLTRANS. Moreover, we also simulated and reconstructed at  $1x10^{12}$  numbers of photons per bunch. The last result showed us that if XFEL performance gold was overestimated. We could obtain the reconstruction images at about twenty time of number of photons per bunch lower than the performance goals of XFEL.

#### CONCLUSION

In this project, we demonstrated the coherent x-ray diffraction imaging system under the XFEL performance goals. We simulated the diffraction pattern of Tomato Bushy Stunt Virus (TBSV) at 1 Angstrom which is atomic scale and  $10^{12}$  photons per bunch according to the expected performance of the upcoming XFEL source. As per the reconstruction image, we can see some structures of the particular single virus molecule compared with the real space image we got from MOLTRANS.

All in all, the simulation showed that Coherent X-Ray diffraction Imaging is practical method to determine the single molecule structure down to angstrom length scales at XFEL.

#### **FURTHER STUDY**

More data set should be brought about; 1 Angstrom X-ray and 5 Angstrom Resolution X ray scattering image should be simulated. Then, work out with the reconstruction in the same way to get the data under photon statistics.

To imitate the real x-ray scattering experiment, noise which would be in scattering pattern should be concerned. Consequently, noise addition in reconstruction image should be required for this study of reconstruction efficiency in this RECON program.