July 2010

Protein Fragment Search Program ver 1.1.1

Developed by:
BioPhysics Laboratory, Faculty of Life and Environmental Science, Shimane University
1060 Nishikawatsu-cho, Matsue-shi, Shimane, 690-8504, Japan
E-mail: aoyagi@life.shimane-u.ac.jp
Web: http://bioinfoenv.shimane-u.ac.jp/aoyagi/english.htm

Overview:

This is a program for searching combinations of amino acid residues from an amino acid sequence based on m/z values. This program is written in Ruby, an open source programming language (http://www.ruby-lang.org/en/). When you open "protein analysis.zip" file the new folder "protein_analysis" will be created. The folder "protein_analysis" contains "config" folder, "programs" folder, "protein_analysis.bat" and "USER_GUIDE.pdf". USER_GUIDE.pdf is written in Japanese. The all information of USER_GUIDE.PDF is described in this instruction.

The ZIP archive "protein analysis.zip" contains the following files:
protein analysis.bat: The program boot file
config: Information of atomic weight and structures of amino acid residues
programs: Program files

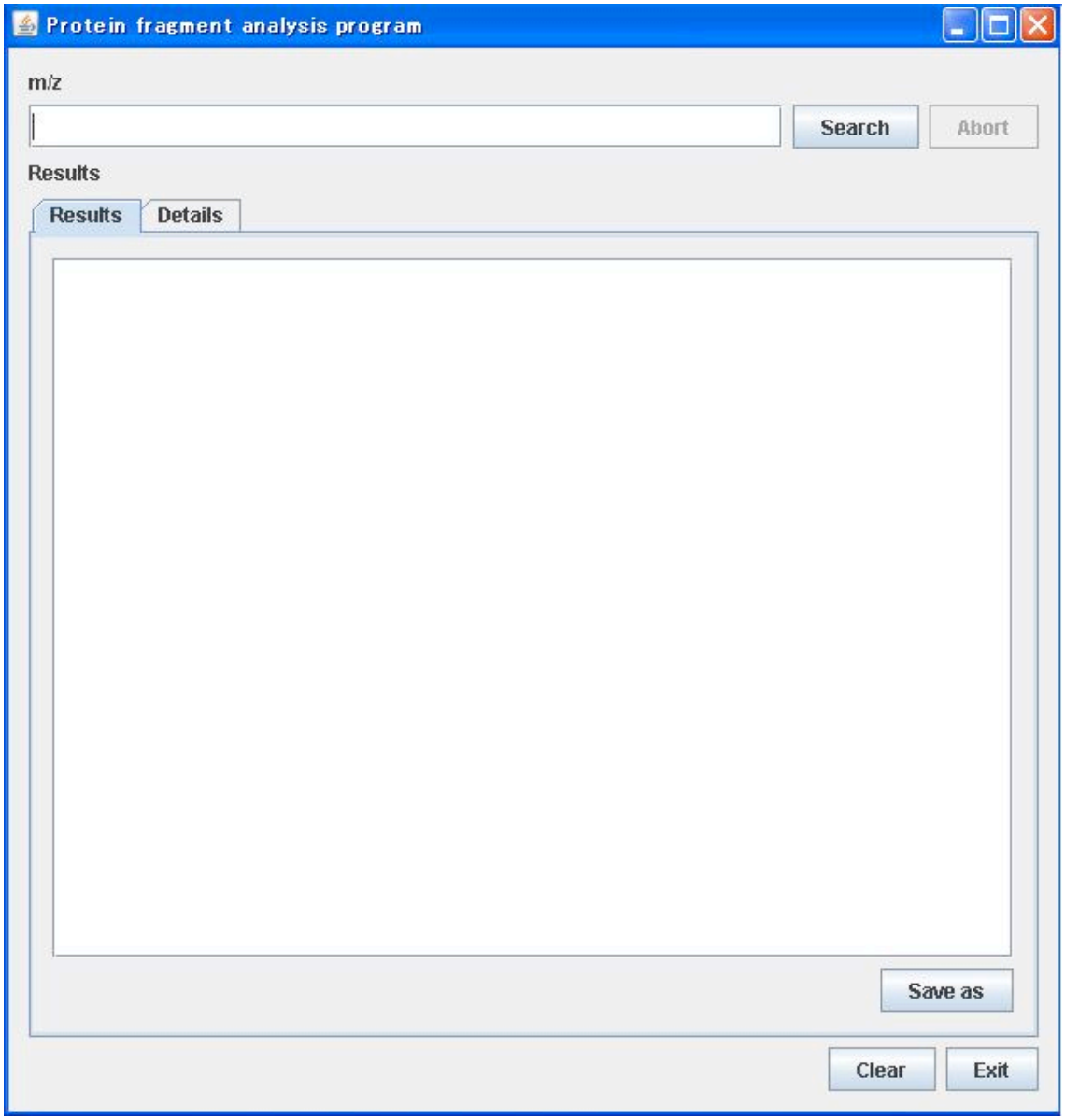
Contents:

Uninstalling:

Please delete "protein_analysis" folder.

How to use:

When you double-click "protein analysis.bat" file, the following window will be opened.



1. Please put m/z value in the "m/z" cell. Put it to two decimal places.
   ex) 70.06
2. Then push "Search" button and then the search of combinations of amino acid residues will begin.
3. "Results" and "Details" windows show the results. "Results" shows combinations of amino acid residues. "Details" shows results of every possible main structure.
4. When you push "Save as", you can save the results as a txt file.

5. When you push "Clear" button, the previous results in "Results" and "Details" windows will be deleted.
6. When you finish the program, please push "Exit" button.

* This program does not consider some of fragments from one amino acid. Table 1 shows fragment ions related to each amino acid reported in the following references.

D. S. Mantus, B. D. Ratner, B. A. Carlson, and J. F. Moulder, Anal. Chem. **65**, 1431 (1993)

J-B. Lhoest, E. Detrait, PvdBd Aguilar and P. Bertrand: J. Biomed. Materials Res. **41**, 95 (1998)

Table 1: Reported fragment ions related to the amino acids

Residues	Formula	*m/z*	Residues	Formula	*m/z*	Residues	Formula	*m/z*
Ala	C2H6N	44	Glu	C4H6NO	84	Ser	C2H6NO	60
Arg	CH3N2	43		C4H8NO2	102		C3H3O2	71
	C2H7N3	73	Gly	CH4N	30	Thr	C4H5O	69
	C4H10N3	100	His	C4H5N2	81		C3H8NO	74
	C4H11N3	101		C4H6N2	82	Trp	C9H8N	130
	C5H10N3	112		C5H8N3	110		C10H11N2	159
	C5H11N4	127	Ile	C5H12N	86		C11H8NO	170
Asn	C3H4NO	70	Leu	C5H12N	86	Tyr	C7H7O	107
	C3H7N2O	87	Lys	C5H10N	84		C8H10NO	136
	C3H6NO2	88	Met	C2H5S	61	Val	C4H10N	72
	C4H4NO2	98	Phe	C8H10N	120		C5H7O	83
Asp	C3H6NO2	88		C9H7O	131			
Cys	CHS	45	Pro	C4H6N	68			
Gln	C4H6NO	84		C4H8N	70			

Results:

The results will be written using abbreviated expressions as follows.

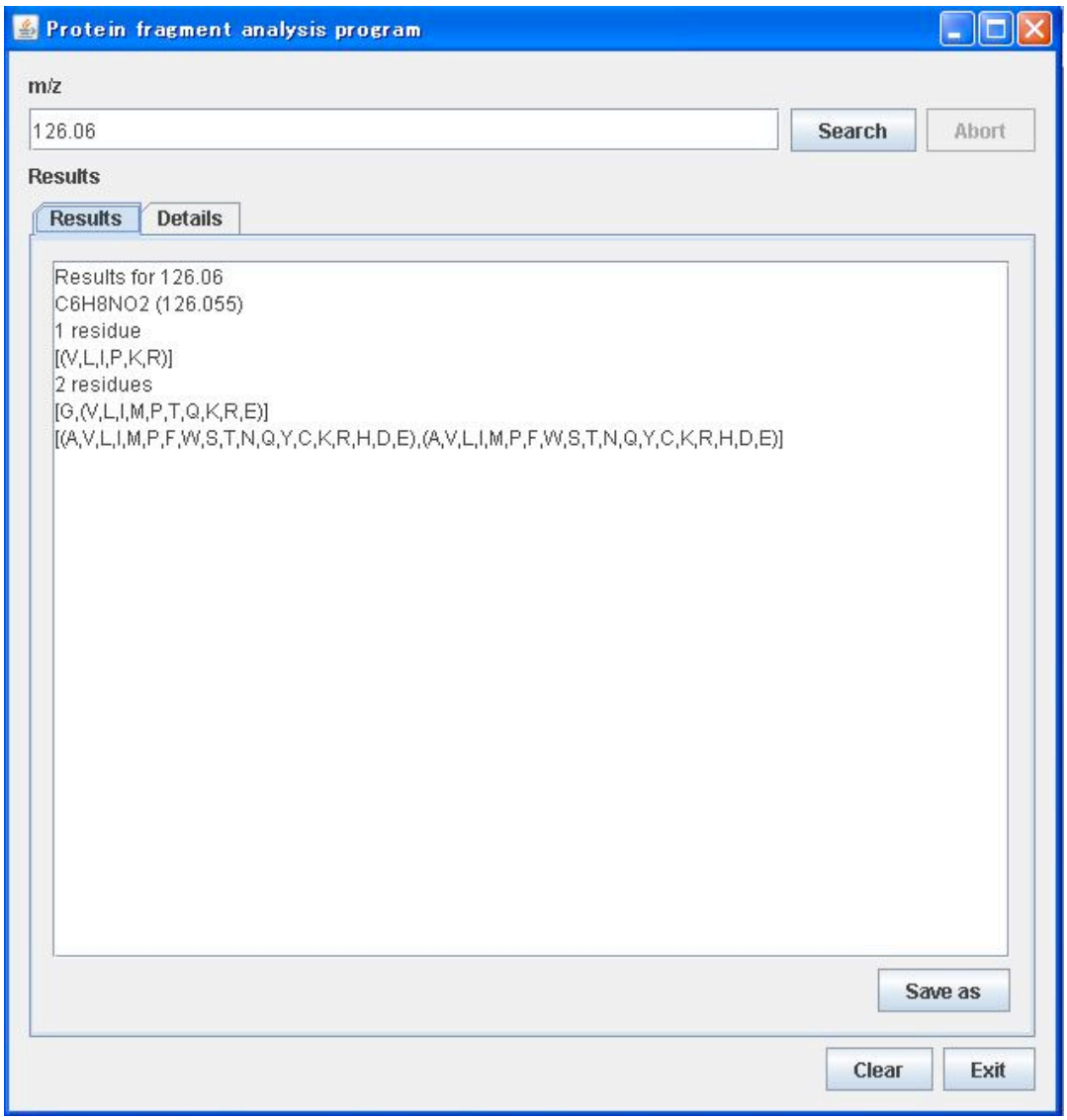
[Y, W] means that a secondary ion can be generated from neighboring Y and W residues.

[(Y, W, T)] means that a secondary ion can be generated from Y, W or T residue.

[G, (A, E)] means that a secondary ion can be generated from neighboring "G and A" or "G and E".

[G, (A, E), (S, T)] means that a secondary ion can be generated from neighboring "G, A and S" or "G, A and T" or "G, E and S" or "G, E and T".

Example: m/z 126.06



The comment on [Results]:

6H8NO2 (126.055) => The possible formula of the secondary ion m/z 126.06 is $C_6H_8NO_2$.

1 residue => When the secondary ion was generated from 1 residue:

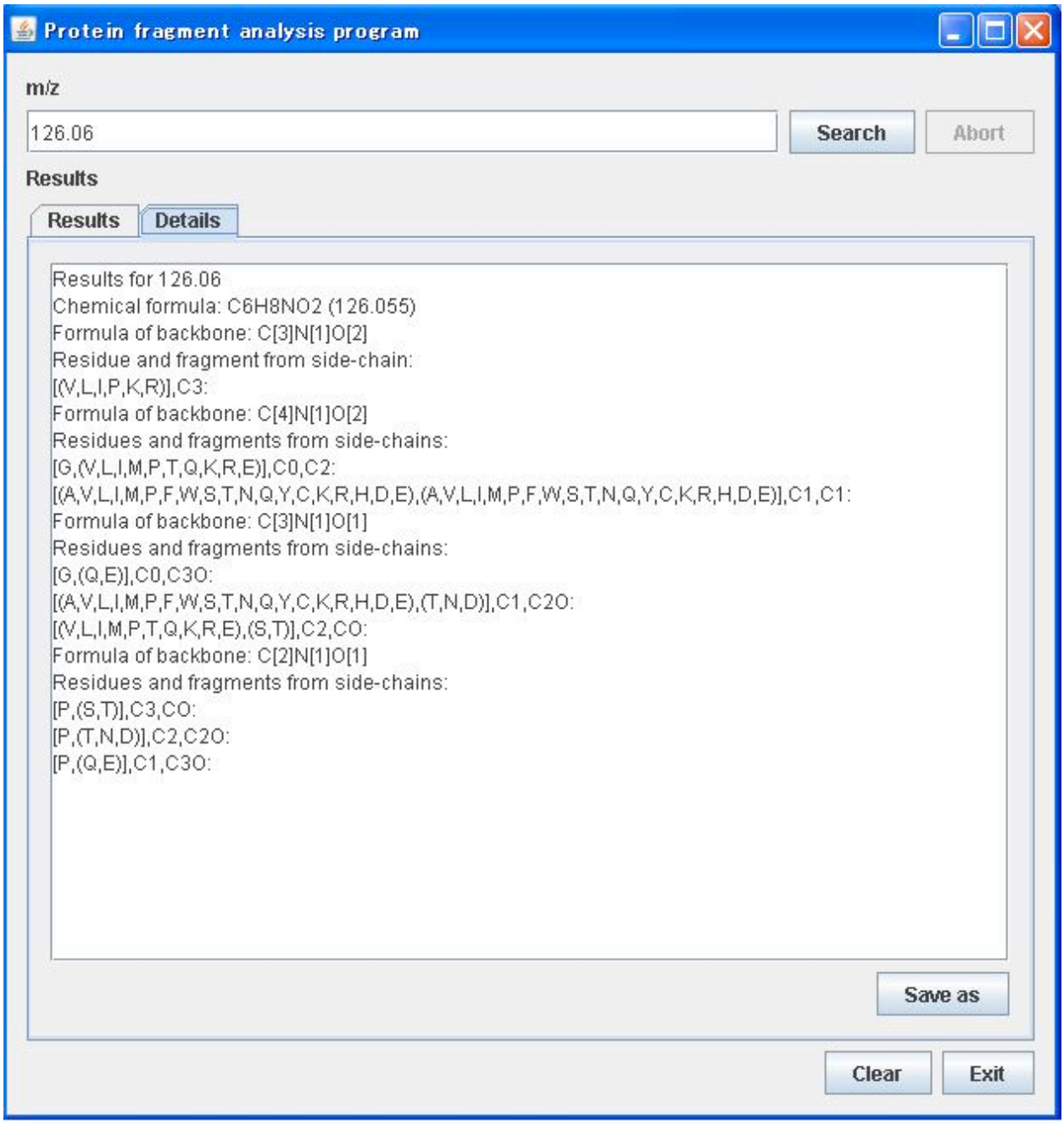
[(V,L,I,P,K,R)] => The secondary ion can be generated from V,L,I,P,K, or R residue.

2 residues => When the secondary ion was generated from 2 residues:

[G,(V,L,I,M,P,T,Q,K,R,E)] => The secondary ion can be generated from two neighboring residues, G and V, G and L, G and I, G and M, G and P, G and T, G and Q, G and K, G and R, or G and E.

[(A,V,L,I,M,P,F,W,S,T,N,Q,Y,C,K,R,H,D,E),(A,V,L,I,M,P,F,W,S,T,N,Q,Y,C,K,R,H,D,E)]

=> The secondary ion can be generated from two neighboring residues, one of "A,V,L,I,M,P,F,W,S,T,N,Q,Y,C,K,R,H,D and E resides" and one of "A,V,L,I,M,P,F,W,S,T,N,Q,Y,C,K,R,H,D and E resides".



The comment on [Details]:

Chemical formula: C6H8NO2 (126.055)

Formula of backbone: C[3]N[1]O[2]

=> The part of the secondary ion related to the backbone of the polypeptide is composed of $C_3H_xNO_2$.

Residue and fragment from side-chain: [(V,L,I,P,K,R)],C3:

=> The part of the secondary ion related to the side chain of the polypeptide is composed of C_3H_x which can be generated from the side chain of V, L, I, P, K or R residue.

Advanced usages:

Addition of elements and modification of atomic weight:

Elements used for the program were defined in "atom weight.yml" file in "config" folder.

One row contains definition of one element like the following:

<Symbol of element><colon><space><atomic weight>

ex)

C: 12

H: 1.0078

Please do not add a space at the beginning of a row. Please back up the original file before you modify this file.

Addition or modification of principal chain structures of amino acid residues:

Possible structures of principal chains of amino acid residues corresponding to the number of residues are described in the following file in "config" folder:

_ principle chains.yml

_ principle chains with proline.yml

"principle chains.yml" contains all amino acids information. "principle chains with proline.yml" contains structures of principal chains of amino acids when more than one proline residue is included. Files "principle chains.yml" and "principle chains with proline.yml" are read when the program is started. If they are changed wrongly, the program will not be performed properly.

Possible fragment formulas from principal chains are described in each file. Up to five neighboring amino acid residues are considered in this program. Fragments from principal chains (backbone) and those from side chains are considered separately. According to the number of parincipal chains composing a secondary ion, every fragment of each side chain is checked if it can be a part of the secondary ion based on its mass value.

For example, possible formulas of fragments are listed from the least number of amino acid residues which the secondary ion can contain. The "principle chains.yml" file contains the following list. First group "- C3N2O2, - C3NO2, - C2N2O, - C2NO, - C2O, - CN, - C" contains possible fragments from a principal chain of one amino acid residue, and second group "- C5N3O3, - C5N2O3, - C4N3O2, -

C4N2O2, - C4NO2, - C3N2O, - C3NO" contains possible fragments from principal chains of two neighboring amino acid residues.

-
- C3N2O2
- C3NO2
- C2N2O
- C2NO
- C2O
- CN
- C
-
- C5N3O3
- C5N2O3
- C4N3O2
- C4N2O2
- C4NO2
- C3N2O
- C3NO

The basic concept of the protein fragment search method:

Fragment ion peaks from neighboring amino acids are identified by searching every combination of amino acids based on the following hypothesis:

1) The number of carbon, oxygen, nitrogen and sulfur atoms is considered.

2) The number of hydrogen atoms is adjusted (hydrogen addition and hydrogen desorption are considered flexibly).

3) Double bonds do not cleave.

4) Recombination and rearrangement of ions are neglected.

Possible fragments from every part of a protein molecule were considered with structures of the 20 amino acid residues composing a protein. Table 2 shows fragments from the side chains of the residues, and table 3 shows fragments from the principal chain or principal chains (backbone). If it is assumed that a fragment ion is generated from a part containing two neighbor residues, its formula can be one of the combinations of two of the residues in table 2 and fragments of the two principle chains in table 3. In addition, when the fragment ion generation part contains a proline residue, the possible fragments in table 4 should be considered.

Table 2 Possible fragments generated from side chains of amino acid residues

Amino acid		Formula of side chain	Possible fragment from side chain (except for hydrogen)					
Gly	G	H						
Ala	A	CH_3	C					
Val	V	C_3H_7	C_3	C_2	C			
Leu	L	C_4H_9	C_4	C_3	C_2	C		
Ile	I	C_4H_9	C_4	C_3	C_2	C		
Met	M	C_3H_7S	C_3S	C_2S	C_2	C		
Pro	P	C_3H_6	C_3	C_2	C			
Phe	F	C_7H_8	C_7	C				
Trp	W	C_9H_7N	C_9N	C				
Ser	S	CH_3O	CO	C				
Thr	T	C_2H_5O	C_2O	C_2	CO	C		
Asn	N	C_2H_4NO	C_2NO	C_2O	C			
Gln	Q	C_3H_6NO	C_3NO	C_3O	C_2	C		
Tyr	Y	C_7H_8O	C_7O	C_7	C			
Cys	C	CH_3S	CS	C				
Lys	K	$C_4H_{11}N$	C_4N	C_4	C_3	C_2	C	
Arg	R	$C_4H_{11}N_3$	C_4N_3	C_4N_2	C_3N	C_3	C_2	C
His	H	$C_4H_5N_2$	C_4N_2	C				
Asp	D	$C_2H_2O_2$	C_2O_2	C_2O	C			
Glu	E	$C_3H_4O_2$	C_3O_2	C_3O	C_2	C		

Table 3 Possible fragments generated from principle chains (backbone) of amino acid residues

Number of residues	Formula of backbone	Possible fragment from backbone (except for hydrogen)						
1	C_2H_2NO	C3N2O2	C3NO2	C2N2O	C2NO	C2O	CN	C
2	$C_4H_4N_2O_2$	C5N3O3	C5N2O3	C4N3O2	C4N2O2	C4NO2	C3N2O	C3NO
3	$C_6H_6N_3O_3$	C7N4O4	C7N3O4	C6N4O3	C6N3O3	C6N2O3	C5N3O2	C5N2O2
4	$C_8H_8N_4O_4$	C9N5O5	C9N4O5	C8N5O4	C8N4O4	C8N3O4	C7N4O3	C7N3O3
5	$C_{10}H_{10}N_5O_5$	C11N6O6	C11N5O6	C10N6O5	C10N5O5	C10N4O5	C9N5O4	C9N4O4

Table 4 Possible fragments generated from principle chains containing Proline

Number of residues	Possible fragment from principle chain (except for hydrogen)		
1	N		
2	C2N2O	C2NO	
3	C4N3O2	C4N2O2	C5N3O3
4	C6N4O3	C6N3O3	C7N4O4
5	C8N5O4	C8N4O4	C9N5O5

Fragments and possible side chains which can generate each fragment:

Zero	G
C	A V L I M F W S T N Q Y C K R H D E
C2	V L I M T N Q K R E
C3	V L I (P) K R
C4	L I K
C7	F Y
CO	S T
C2O	T N D
C2O2	D
C2NO	N
C3O	Q E
C3O2	E
C3NO	Q
C3N	R
CS	C
C2S	M
C3S	M
C4N	K
C4N2	R H
C4N3	R
C7O	Y
C9N	W

References:

More information on “Protein Fragment Search Program” is available on the following web site.

http://bioinfoenv.shimane-u.ac.jp/aoyagi/english.htm

Related articles:

• S. Aoyagi, Surface and Interface Analysis, 41(2) 136-142, (2009)

• S. Aoyagi, M. Higuchi, N. Kato, and M. Kudo, e-J. Surf. Sci. Nanotech., 7, 715-720 (2009)

If you have any questions on this program, please contact us.

e-mail: aoyagi@life.shimane-u.ac.jp