Structure Determination Tutorial

Michael Baran
G.V.T. Swapna
Jim Aramini
Roberto Tejero
Janet Huang
Nadeem Riaz

6/26/03
7/28/03

Contents

I. Process Backbone Experiments using standard AutoProc parameters

II. Peak Picking

    II.A Identify NH2 Residues in HSQC

    II.B Identify folded peaks from large sw HSQC

    III.B Peak Pick Triple Resonance Experiments

III. Backbone Assignments

IV. Side Chain Assignments

    IV.A Aliphatic Sidechain assignments

    V.B Assign NH2's

    IV.C Aromatic Side Chain Assignments

    IV.D Run missing_shifts.pl

V. Structure Determination

    V.A. AutoStructure

    V.B Run Talos for dihedral angle constraints

    V.C Analyze HNHA experiments to determine J values

    V.D Xplor Refinement

I.A Process Backbone Experiments using standard AutoProc parameters

Process each experiments xy and xz plane to set correct phasing.

Use proper zero filling. CABM data collection usually use 1024 x 512 x 256 matrix. For C13-edited NOESY use 1024x512x512

Use | nmrPipe –fn EXT –left –sw flag for 3d processing where appropriate. This command will exclude half of spectra where no peaks appear and thus reduce processing time.

II. Peak Picking (Triple Resonance Experiments)

II.A Identify NH2 Residues in HSQC

1. Open HSQC experiment.
    1. Display what nuclei is on each axis
        i. Sparky cmd: xa
    2. Rotate cube so x1 is H and x2 is N
        i. Sparky cmd: xx and xr
    3. Set proper contour levels
        i. Sparky cmd: ct
    4. Adjust aspect ratio
        i. Sparky cmd: vt

2. Overlay NH2-HSQC onto NH-HSQC
    1. Sparky command – ol

3. Pick all HSQC peaks
    1. Sparky cmd: F11 and draw box around all peaks to select

4. Remove NH2 peaks from HSQC
    1. Select peaks with F1 and delete from list

5. Run examine_expected_peaks.pl to be sure all peaks are accounted for OR manually determine how many peaks expected
    1. Num peaks = num or residues - # of PRO – 1 terminal NH?

II.B Identify folded peaks from large sw HSQC

III.B Peak Pick Triple Resonance Experiments

1. Peak pick HNCO, HNCACB, HNcoCACB and any other triple resonance experiments against HSQC.

2. Open triple resonance experiment as described in II.A.1

3. Synchronize spectra to HSQC
    a. Sparky cmd: yt
    b. Synchronize each nuclei individually
4. Restrictive peak pick spectra
    a. Sparky cmd: kr
    b. use default tolerances
    c. adjust contour levels (ct) as needed until roughly expected number of peaks are picked
        i. CO (intra) experiments – expect approx same number of peaks as residues
        ii. no CO (inter) experiments – expect 2X as many peaks as residues
5. Convert peak lists to autoassign format and perform peak list validations
    a. run extract_columns.pl – to extract chemical shift columns and data height
    b. run create_peak_list.pl – to convert extract_colums output to autoassign format lists
6. Run examine_expected_peaks on each spectra to evaluate if number of peaks picked is reasonable
7. Run calculate_registration.pl – large registration errors can be used to detect flawed peak lists.
    a. register each spectra to HSQC to register N H dimensions
    b. register C dimension by registering spectra to HNcoCACB
8. Run examine_spin_systems.pl report. ESS report can be used to detect incomplete or overlapped spin systems.

III. Backbone Assignments

1. Convert peak lists to AutoAssign format as described above in step II.B.5.
2. Register peak lists using calculate_registration.pl as described in step III.B.7
    a. Generate strategy specific peak list (autoassign runs better with HNCA and HNcoCA data. However, if these experiments are absent they can be "created" from HNCACB and HNcoCACB experiments
        i. Ex. Gen HNCA from HNCACB
        ii. Ex2. Generate HNcoCA from HNcoCACB
        iii. Use extract_by_phase and extract_by_filter.pl to produce these "pseudo-peak lists"
    b. Use registration value to shift peaks lists using apply_shifting.pl
    c. Calculate AutoAssign tolerances
        i. Take largest Standard deviation and apply by 3 or 4 for each dimension
        ii. Example of typical tolerances are: N: 0.01 H: 0.11 CA: 0.21

CB: 0.21 CO: 0.4 HA: 0.02

3. Write AutoAssign Control file
4. Write AutoAssign Override file
    a. overrides from HCCH-TOCSY
    b. Gly phased HA experiment overrides
    c. Selective perdueteration overrides
5. Run AutoAssign to make assignments
    a. Save cmap files/images of assignments
    b. Save bmrb file of assignments
6. Run validate_assignments.pl to evaluate problematic assignments
7. Paste assignments back into spectra for manual validation
    a. Bmrb2SparkyPks.pl - script to convert autoassign output bmrb file to a peak list which can be opened in Sparky.
    b. Open sparky HSQC
    c. Read Assigned Sparky List from Bmrb2SparkyPks.pl into each peak list
        i. Sparky cmd: rp
    d. Synchronize Peak lists and verify assignments.
    e. Edit peak lists and re-run autoassign until satisfied.

IV. Side Chain Assignments

IV.A Aliphatic Sidechain assignments

1. Process HCCH-TOCSY and hCCH-TOCSY
2. Use these spectra with triple resonance backbone experiments and average carbon and hydrogen side shift values to tentatively assign all spin system Carbons and Hydrogens.
3. Process HCCH-COSY
4. Make HH and CH projection of HCCH-COSY
    a. Sparky cmd: (XH) ucsfdata –s1 512 –r –o [out] [in]
    b. Sparky cmd: (HH) ucsfdata –s1 256 –r –o [out] [in]
    c.
5. Use projections to restrictive peak HCCH-COSY
6. Run 1 cycle of AutoStructure to assign HCCH-COSY
7. Paste AutoStructure assignments into Sparky HCCH-COSY
    a. Sparky cmd: rp
8. Use AutoStructure QM file to help find un-assigned residues
9. Manually confirm side chain assignments using both HCCH-TOCSY ,hCCH-TOCSY and HNCACB.
    a. Use expected H chemical shift chart as well as expected peak pattern knowledge as a guide.
    b. Frequently view assignments standard deviations using the sparky rl command. Keep all standard deviations below 0.1.
10. Validate side chain assignments using validate_assignments.pl
    a. Save rl results from sparky

b. Use createBmrb.pl to generate bmrb file used as input for validate_resonace_Assignemnts.pl

IV.B Assign NH2's

I would suggest assigning NH's after running an initial AutoStructure run. This way the assignments from AutoStructure can be pasted into the NOESY to locate the given residue strip.

Use NOESY's to stereo-specifically assign NH2's. based on peak intensities and expected peak positions.

IV.C Aromatic Side Chain Assignments

Again I would suggest that aromatic assignments are not made until an initial AutoStructure run has been completed. This way assignments can be pasted into the C13-edited NOESY and N15-edited NOESY.

Process Aromatic C13-edited HSQC and aromatic NOESY.
Use these experiments along with aliphatic NOESY's to assign aromatics.

(If using Yamasaki RD experiments good luck!)

IV.D Run missing_shifts.pl

Run AutoPeak missing_shifts.pl to evaluate assignment completeness.

V. Structure Determination

V.A. AutoStructure

NOTE: use –j flag to include jcoupling data
HYPER and Talos may conflict w/ each other, use caution when running both

1. Process N15-edited NOESY and C13 edited NOESY
    a. Make sure to use LP!! Process C13-NOESY w/ 1024x512x512 matrix
2. Peak pick noesy's against respective HSQC as well as HH projection.
3. Generate AutoStructure Input
4. Run auto-structure
5. analyze AutoStructure QM file for missed residues in peak lists
6. Re-run auto structure until satisfied with quality scores.
    a. M score < 0.1 N15 NOESY
    b. M score < 0.4 C13 NOESY
7. Then run multiple cycles of AS analyzing the unassigned file.

8. Remove spurious peaks which are being unassigned and re-run AS until ~under 10 assignments are being removed per cycle.
9. Calculate I,L scores

V.B Run Talos for dihedral angle constraints

1. run Tab4Talos.pl script to convert bmrb assignments into Talos intput
2. run talos
3. run Talos2Dyana.pl to convert talos output to Dyana format for use in AutoStructure

V.C Analyze HNHA experiments to determine J values

NOTE: Jvalues 6-8 sheet
2-4 values

1. peak pick HNHA
2. In order to calculate J values need to separate diagonal and off-diagonal peaks
    a. **OFF-DIAG peaks**
    b. Bmrb2SparkyPks.pl HA N H
    c. Paste rp
    d. Bmrb2SparkyPks.pl HA2 N H
    e. Paste rp
    f. Bmrb2SparkyPks.pl HA3 N H
    g. Paste rp
    h. DIAGONAL PEAKS
    i. Bmrb2SparkyPks.pl H N H
    j. Paste rp
3. Create SpreadSheet in Excel
4. columns
    a. Diag Assignment
    b. Diag peak intensity
    c. Off- Diag peak assignment
    d. Off-Diag peak intensity
    e. 1.1*I(x)/I(d)
    f. sqrt(1.1*I(x)/I(d))
    g. atan(sqrt(1.1*I(x)/I(d)))
    h. atan(sqrt(1.1*I(x)/I(d))) / 2 * pie * 0.0125
        i. execel pie = 3.14 cladiograph pie = 180
5. Convert to bmrb format manually

assign	diag	off-diag ass	off-diag I				
I3H-N-H	405877	I3HA-N-H	-146432	-0.396857176	0.629966012	0.562162412	7.161304616
I4H-N-H	164581	I4HA-N-H	-125338	-0.837713952	0.915267148	0.741186245	9.441862988
S5H-N-H	823571	S5HA-N-H	-161818	-0.2161317	0.464899666	0.435175174	5.543632785

I6H-N-H	564057	I6HA-N-H	-301423	-0.58782233	0.766695722	0.654101023	8.332497114
S7H-N-H	351641	S7HA-N-H	-211607	-0.661946986	0.813601245	0.682979536	8.700376249
T9H-N-H	428115	T9HA-N-H	-273439	-0.702575009	0.838197476	0.697602059	8.886650437
N11H-N-H	546969	N11HA-N-H	-346288	-0.696413874	0.834514154	0.69543475	8.859041395
N13H-N-H	347744	N13HA-N-H	-151580	-0.479484908	0.692448488	0.605639835	7.715157136
T14H-N-H	369613	T14HA-N-H	-213263	-0.634688986	0.796673701	0.672709417	8.569546719
M15H-N-H	246989	M15HA-N-H	-191134	-0.851241958	0.922627746	0.745176895	9.492699302
K16H-N-H	315093	K16HA-N-H	-175804	-0.613737531	0.783414023	0.664545403	8.465546532
I17H-N-H	168571	I17HA-N-H	-103481	-0.67525909	0.821741498	0.687858078	8.762523283
T18H-N-H	314057	T18HA-N-H	-152424	-0.533872514	0.730665802	0.631011961	8.038368926
L19H-N-H	167629	L19HA-N-H	-107724	-0.706896778	0.840771537	0.699112008	8.905885455
S20H-N-H	454340	S20HA-N-H	-55344	-0.133993045	0.366050604	0.350901634	4.470084507

J BMRB format for AutoStrucutre

1	3	JHNHA	3	ILE	H	3	ILE	HA	7.161305	1.5
2	3	JHNHA	4	ILE	H	4	ILE	HA	9.441863	1.5
3	3	JHNHA	5	SER	H	5	SER	HA	5.543633	1.5
4	3	JHNHA	6	ILE	H	6	ILE	HA	8.332497	1.5
5	3	JHNHA	7	SER	H	7	SER	HA	8.700376	1.5
6	3	JHNHA	9	THR	H	9	THR	HA	8.88665	1.5
7	3	JHNHA	11	ASN	H	11	ASN	HA	8.859041	1.5
8	3	JHNHA	13	ASN	H	13	ASN	HA	7.715157	1.5
9	3	JHNHA	14	THR	H	14	THR	HA	8.569547	1.5
10	3	JHNHA	15	MET	H	15	MET	HA	9.492699	1.5
11	3	JHNHA	16	LYS	H	16	LYS	HA	8.465547	1.5
12	3	JHNHA	17	ILE	H	17	ILE	HA	8.762523	1.5
13	3	JHNHA	18	THR	H	18	THR	HA	8.038369	1.5
14	3	JHNHA	19	LEU	H	19	LEU	HA	8.905885	1.5
15	3	JHNHA	20	SER	H	20	SER	HA	4.470085	1.5
16	3	JHNHA	21	GLU	H	21	GLU	HA	8.226777	1.5

V.D Xplor Refinement

1. Use following script:
    a. ShifttoXplor <assignment.bmrb> [output]
2. Use Xplor constraints from AutoStructure
    a. Read xplor constraints into PdbStat and rename nomenclature conflicts
        i. Rea pdb XX.pdb
        ii. Rea cons xplor XX.xpl
            1. find bad Me-Me contacts
            2. change all CZ to HE#
            3. change all CG to HD#
3. Convert DYANA aco file to Xplor w/ PdbStat (may want to remove HYPER constraints and keep only Talos constraints before conversion to Xplor format)
    a. Rea aco xplor XX_dihe.tbl
4. Convert Hbonds to Xplor format w/ PdbStat

    a. rea upl dyana hbond.upl
    b. rea lol dyana hbond.lol
    c. write cons xplor XX_hbond.tbl
5. generate topology file
    a. /farm/software/xplor-nih-2 <generate_psf
6. make pdb for Xplor
    a. rea coo pdb XX.pdb
    b. write coor pdb XX_0.pdb
7. ./RefineProc XX 10 10 PBS 0
8. Make PDB file of 10 best structures
    a. Concatenate all 100 pdb files
    b. Agrupa zr18_sa*.pdb
    c. Pdbstat
    d. Rea coors all
    e. Classify
    f. Del 11 100
    g. Order 0.9
    h. Rmsd best backbone
    i. Write coor pdb best.pdb all

NOTES on getting Xplor to work:

1. Be sure you are added as a user to PBS with proper permissions
2. be sure paths are set correctly for DQS or PBS in RefineProc script
3. set the following paths
    a. export TOPPAR=/farm/software/xplor-nih-2.0.6/toppar
    b. export PATH=/farm/software/bin:${PATH}
    c. export C13SHIFTS=/farm/software/xplor-nih-2.0.6/databases/c13shifts
    d. export GAUSSIANS=/farm/software/xplor-nih-2.0.6/databases/torsions_Gaussians

Ribbons

GENRAL EXECUTION

Creating a Protein structure info

1. run ribbons-model <pdb> <model name> Produces ss structure info
2. Run atom-model <pdb> Produces atom sphere and bond info
3. To create disulfide bond info, go to File->Export->SG Bonds
4. Create ribbon to side chain bond info File->Export->CG Bonds
5. Polish CG bonds so only have active compoenent parts

Coloring Secondary Structures

Add another tag to

Important programs

pdb-range-ss: allows for coloring certain ss different colors than default
pdb-sele-pdb: allow to cut certain residues outa pdb file

MODEL CREATION

Steps:

1. get PDB
2. Created base model structure (ribbons-model <pdb file>)
3. Create atom model (atom-model <pdb file>)
4. Correct secondary structure to match pdb file
5. If certain important functional sites, create pdb file w/only those residues

7. run atom-model on aromaitc pdb file
8. run pdb-res-sph with color.1 color file to get color coded sphere
9. run sph-bond on file created in 8 to get color coded cylinders
(coded by residue) * Note 8 & 9 need to be rerun to
change colors of cylinders *
10. Done!
11. correct _cg_bonds.cyl file

For texts, see RBFA for example .text file

RIBBONS CODES

Ribbons Coloring Codes:

#	Color
1	Red
2	Green
3	Yellow
4	Blue
5	Magenta
6	Cyan
7	White
8	Orange
9	Grey
10	Lapis
11	Silver
12	Lavendar
13	Brown
14	Gold
15	Rose
16	Lightgreen

Ribbons Secondary Structure Codes

Letter	Structure

S Sheet
A Sheet w/Arrow
H Right handed alpha helices
L Left handed alpha helices
3 3-10 helices
c coil

RIBBONS FILE INFO

.ribbons file contains names of polypeptide ss files and optional color file

.ss file contain polypeptide ss info
First line is arbitrary title line
Second line contains coloring keys, 3 or 4 manadatory keys, rest are user definable

.colors file contain color info
See Ribbons-CODES for color code info

.sph contains info on set of atoms (i.e. geometrical representation)

.cyl contains info on set of bonds (i.e. geometrical representation)

.atoms contains list of .sph files (so you can specificall alter one set of atoms)

.bonds contains list of .cyl files

_sg.cyl disulfide bond info

_cb.cyl create bonds from side chains to ribbons
the name field (last field) contains residue info, use that to determine bonds from ribbons to keep and which to dump
p

PYMOL

Very simple to use.

1. Be sure to add secondary structure information to header of PDB file
    a. /farm/software/structure_validation/get_sse_info.pl
    b. cut and paste results of this script into header of pdb
    c. open pdb file with pymol and have fun.

GRASP

The following is how to open your molecule, create a charged surface and create 2 views front and side for publication using grasp. Use the right mouse button to bring up menu as well as select

Calculate charge of molecule:
Read -> Radius/charge File (+assign) -> full.crg

Build Surface:
Build -> Molec Surface -> all atoms

Calculate potential:
Calculate -> New Potential Map -> all atoms
Calculate -> Pot. Via Map and Surface -> Atoms -> all atoms -> all surface
Then adjust values on top bar to increase/decrease redness

To Duplicate images (for front and side picture)
Display -> stereo/split on -> dials both

To bring up backbone wire view:
You must build the backbone wire view. (sorry no command here) Then to see through the potential surface
Panel -> objects -> surface -> transparent