Structure calculation with ARIA and CNS

Seán O'Donoghue, Jens Linge, Wolfgang Rieping, Michael Habeck,
Benjamin Bardiaux, Aymeric Bernard, Thérèse Malliavin, Michael Nilges

€€: EU, Pasteur Institute, CNRS, ...

INSTITUT PASTEUR

Overview

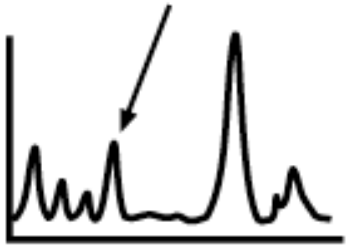
(1) Some basic concepts and history

(2) Recent Developments

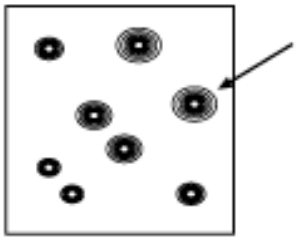
(3) Practical issues

NMR structure determination times

- NMR experiment: weeks to months



- Resonance assignment: weeks to months



- Structural restraints,
- NOE assignment: months

- Structure calculation: fast
- Structure validation: fast

- NOE: basis for high resulution NMR structure
- Interpretation of NOESY spectra complex data analysis problem:

- (1) Assignment and selection of NOEs
- (2) Calculation of structures consistent with data

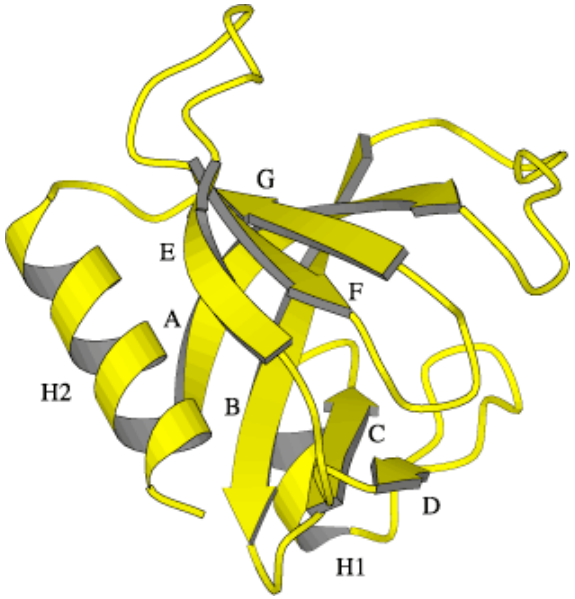
Historical overview over programs

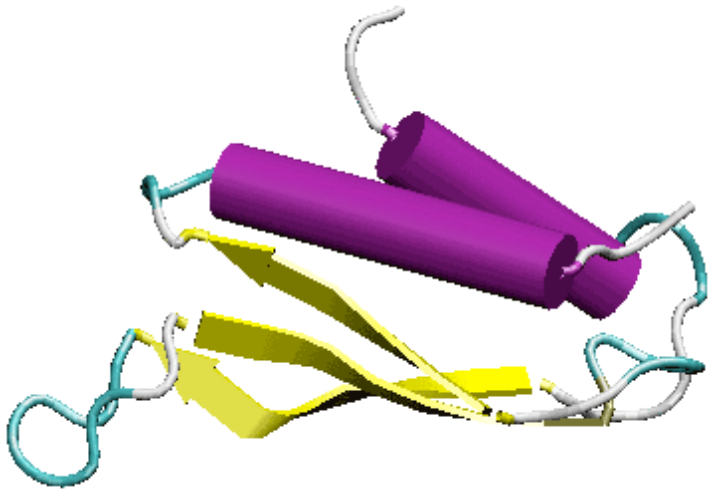
- ASNO
  - automated calculation of potential assignments from a structure
  - several assignments can be stored in XEASY to be picked from manually
    - P Güntert, KD Berndt, K Wüthrich (1993). J. Biomol. NMR, 3: 601–606.
- NOAH (self-correcting distance geometry, SCDG)
  - separate restraint for every assignment possibility
  - identification of incorrect restraints by SCDG
    - C Mumenthaler & W Braun (1995). J. Mol. Biol. 254, 465–480.

- ARIA
  - ambiguities treated by ambiguous distance restraint
  - incorrect restraints recognized by SCDG
  - network anchoring since 2008
    - Nilges et al. (1997) J. Mol. Biol. 269, 408–422.
- CANDID
  - Network anchoring
  - Ambiguous distance restraints
  - SCDG to identify incorrect restraints
    - T Herrmann, P Güntert, K Wüthrich (2002) J. Mol. Biol. 319, 209–22

- AUTOSTRUCTURE
  - "bottom up" construction of assignments by application of expert rules
  - similarities to network anchoring
    - Huang YJ, Tejero R, Powers R, Montelione GT. Proteins: 2006;62:587–603
- PASD
  - more sophisticated version of SCDG
    - J Kuszewski et al (2004). JACS126:6258–6273

ARIA 1994...





- Structures of ß-spectrin PH domain and E. coli dsRBD domain solved with prototype versions of ARIA 1994

Common problems

- chemical shift assignment virtually complete
- ß-sheet topology known
- NOE assignment and manual 3D structure determination failed

- time pressure...

Ambiguous distance restraint (ADR)

- Top-down approach to assign ambiguous NOEs: ADR
  - combine alternate assignment possibilities into single restraint
  - basis: NOE volume depends approximately on distance:

$$V_{ij} \approx \alpha d_{ij}^{-6}$$

  - if calibration constant is the same for all ij, ambiguous V can be converted to estimated target distance D:

$$D \approx \left(\alpha^{-1} \sum_{i=1}^{n_c} V_i \right)^{-\frac{1}{6}}$$

ADR

- ADR first concept that allowed automated NOESY assignment
- First used in calculation of oligomers
  - (Nilges & Brünger, 1991; Nilges, 1993)
- Used in several programs:
  - ARIA–CNS, CAR, CANDID, CYANA
- Treatment of different data types:
  - NOEs, hydrogen bonds, disulfide bridges, paramagnetic shift broadening, chemical shift differences (O'Donoghue and Nilges, 1998)

Iterative structure calculation

- In each iteration, ARIA
  - reduces ambiguity of NOEs
  - detects of inconsistent peaks
- by an analysis of the previous structure ensemble:
  - recalibration of NOEs to distance restraints
  - statistical analysis of restraint violations
  - removal of most unlikely assignment possibilities
- and then recalculates structures

Analysis of structure ensembles

- Use total energy to select a converged structure ensemble
- Calibrate with ensemble averaged theoretical volumes V^{th}

$$\alpha = \frac{\sum_i V_i^{exp}}{\sum_i V_i^{th}}$$

  - using a cutoff V corresponding to cutoff distance (6 Å)
  - theoretical volumes for now from ISPA or relaxation matrix
  - every spectrum is calibrated separately
  - special treatment of first structure calculation round

Violation analysis

- Correct restraints are structurally consistent
- Therefore, incorrect restraints will go against "general trend" imposed by entire data set
- Given restraint is violated if it lies outside bounds by more than user-defined threshold
- Restraint *i* is systematically violated if *f_i* exceeds 0.5
  - (similar to NOAH, CANDID, CYANA,...)

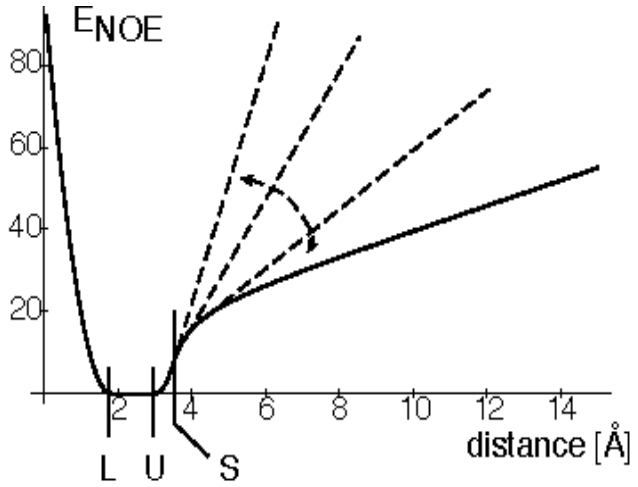
$$f_i = S^{-1} \sum_{j=1}^{S} [\Theta(L_i - d_i^{(j)} + t) + \Theta(d_i^{(j)} - U_i - t)]$$

Mumenthaler, C. and Braun, W. (1995) J. Mol. Biol. 254, 465–480

Error-tolerant structure calculation

- Error-tolerant potentials (NOAH, ARIA-CNS, PASD)
- Violation confinement (CYANA)
- Restraint combination (CYANA)

ARIA-CNS: soft FBHW potential



$$E_{NOE} = \begin{cases} (r - L)^2 & \text{if } r < L \\ 0 & \text{if } L < r < U \\ (U - r)^2 & \text{if } U < r < S \\ A(r - U)^{-1} + B(r - U) + C & \text{if } r > S \end{cases}$$

finite force for large violations

important for automated removal of noise peaks (ARIA)

same as "violation confinement" in CYANA

similar effect as "restraint combination"

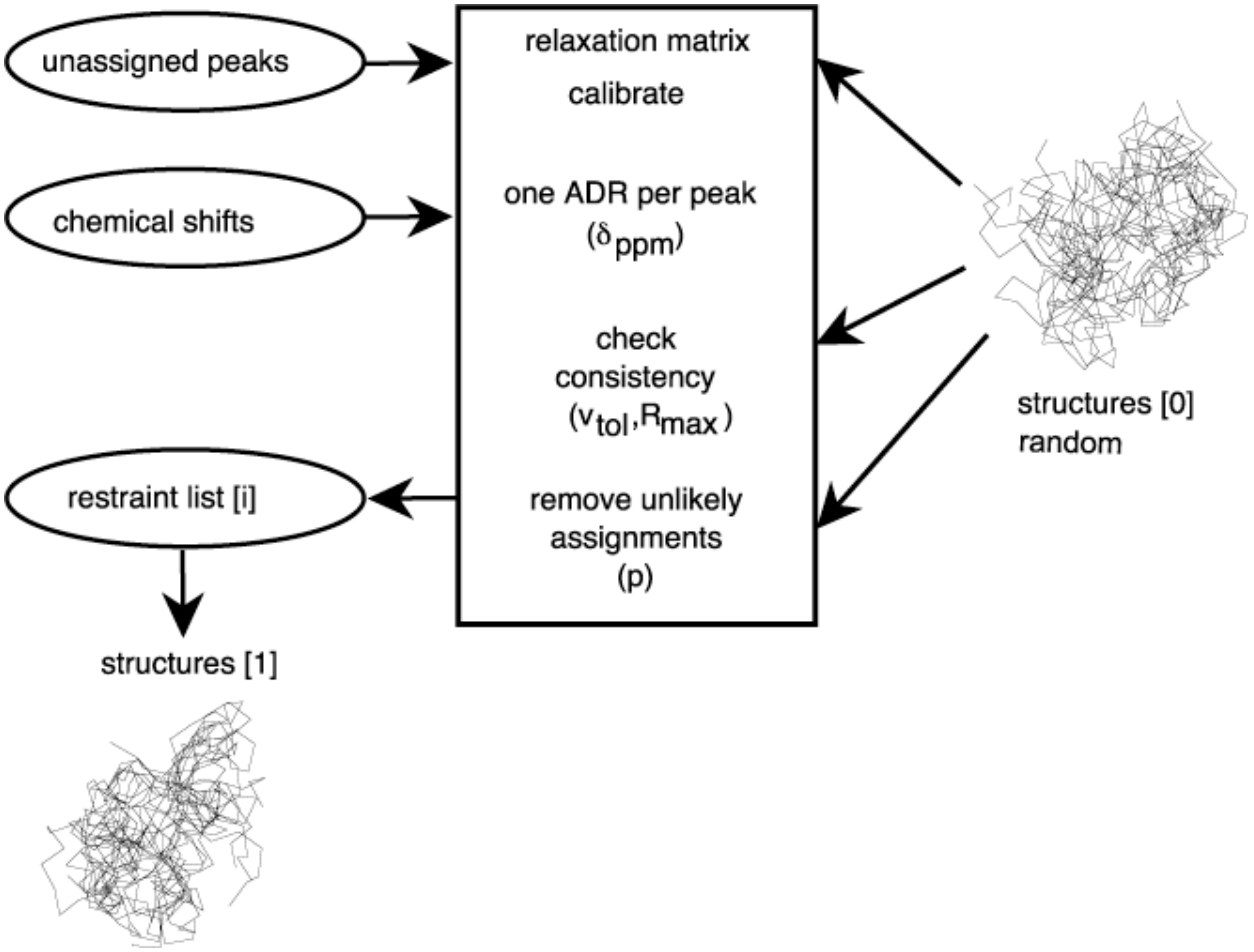
Reduction of assigment possibilities

- For each satisfied ADR:
- calculate partial theoretical volume for each assignment possibility
- order them according to size
- normalize
- remove the smallest possibilities such that the sum of the remaining possibilities exceeds threshold
- decrease this threashold from 1.0 to 0.8

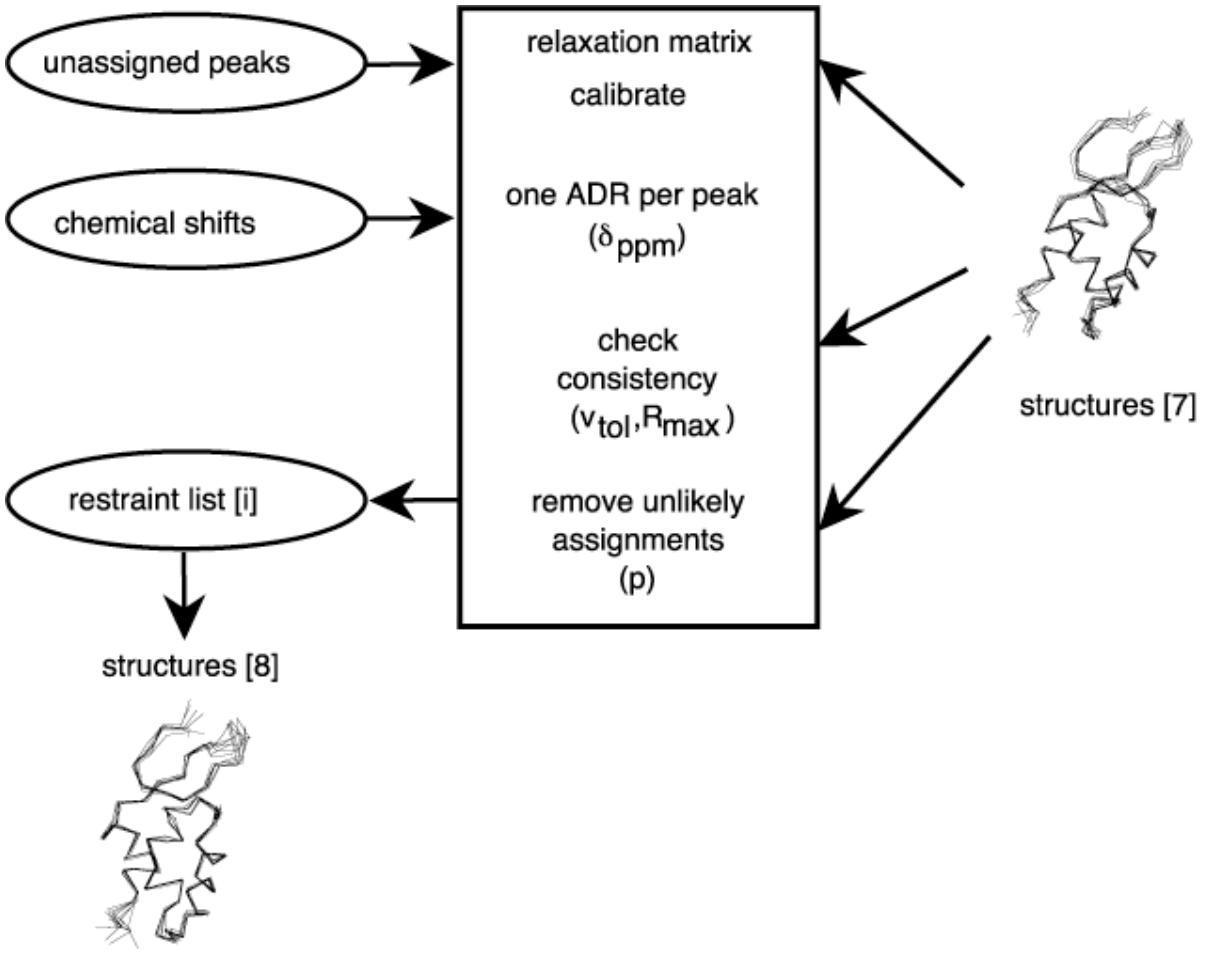
Floating chirality assignment

- Test both possibilities of random selection of prochiral groups at regular intervals
- Keep always lower energy assignment
- Results are stored in ".float" file for each structure

Iterative interpretation of NOE spectra

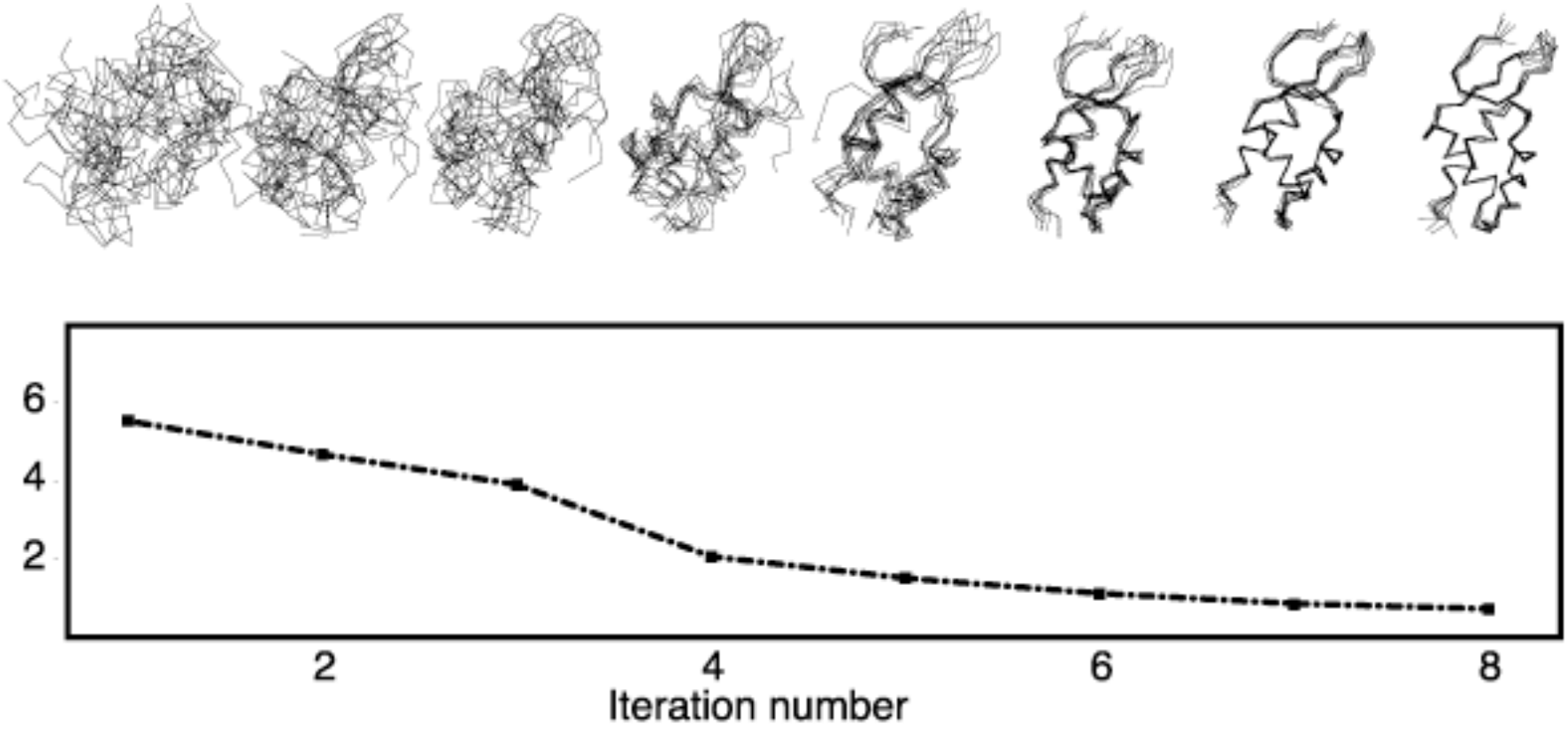


Iterative interpretation of NOE spectra

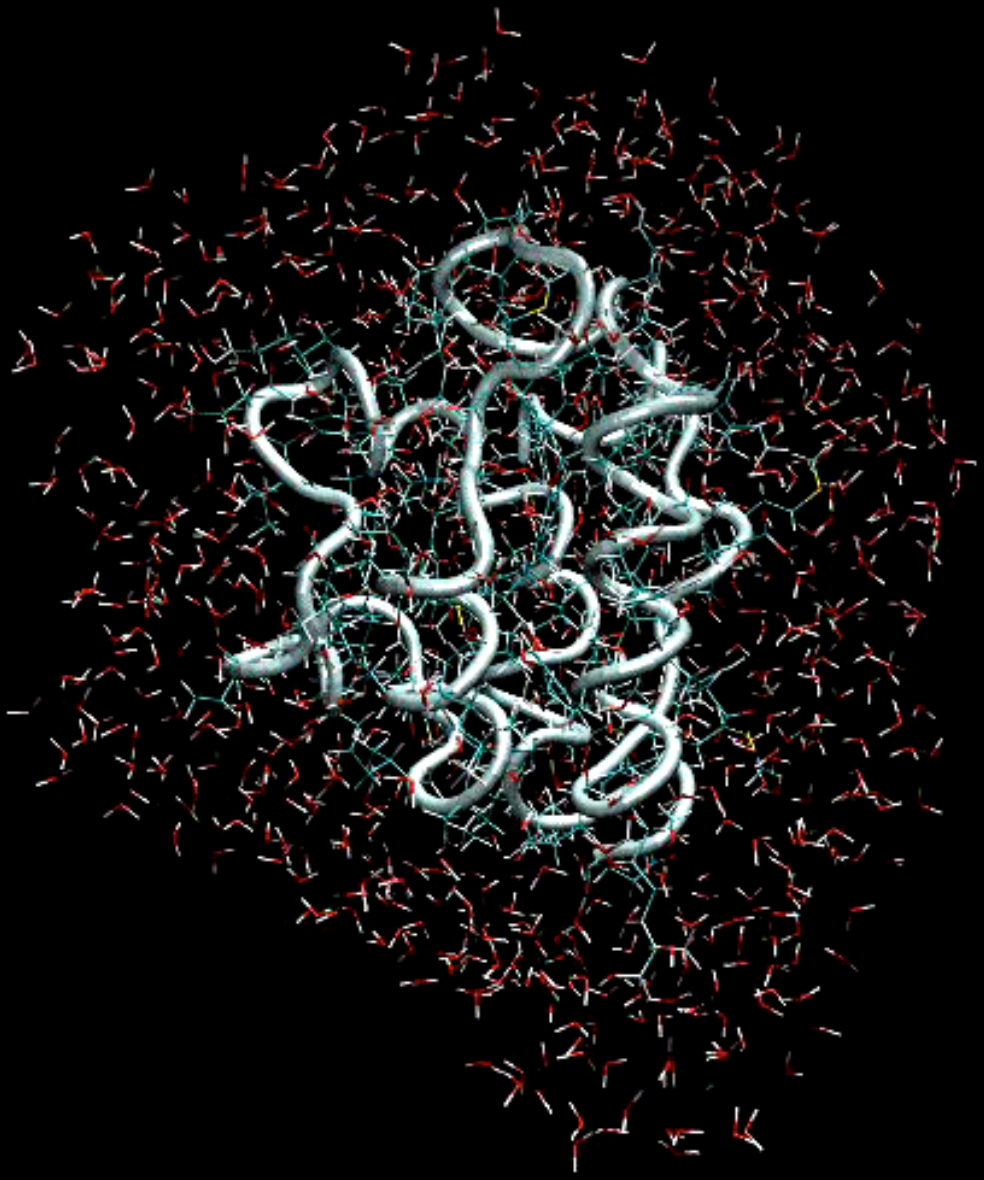


Example convergence

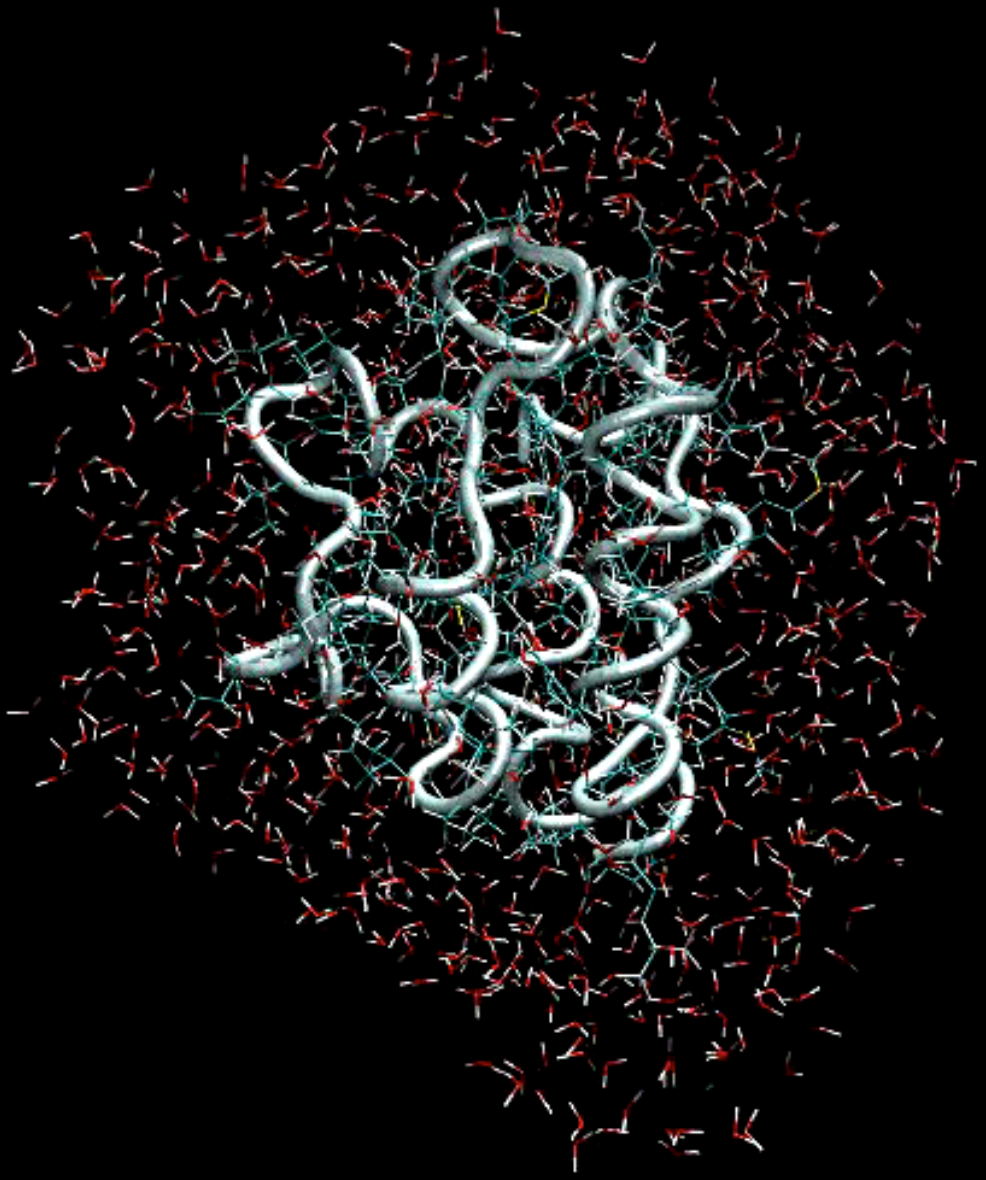
- Completely automated structure calculation and NOESY assignment from noisy ambiguous data (Sunnerhagen et al., 1996)



Final refinement in H2O

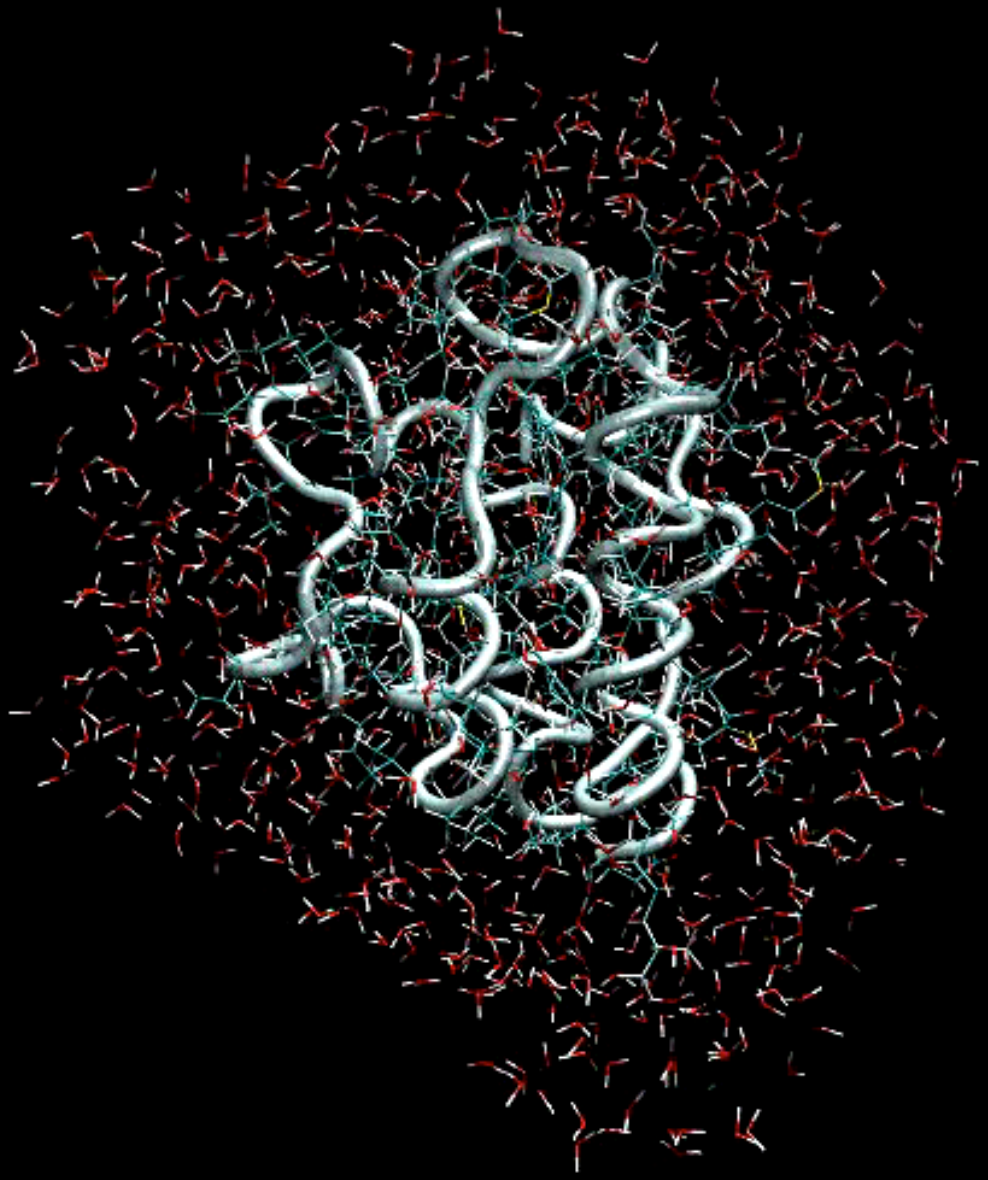


Final refinement in H2O



Automatic in ARIA

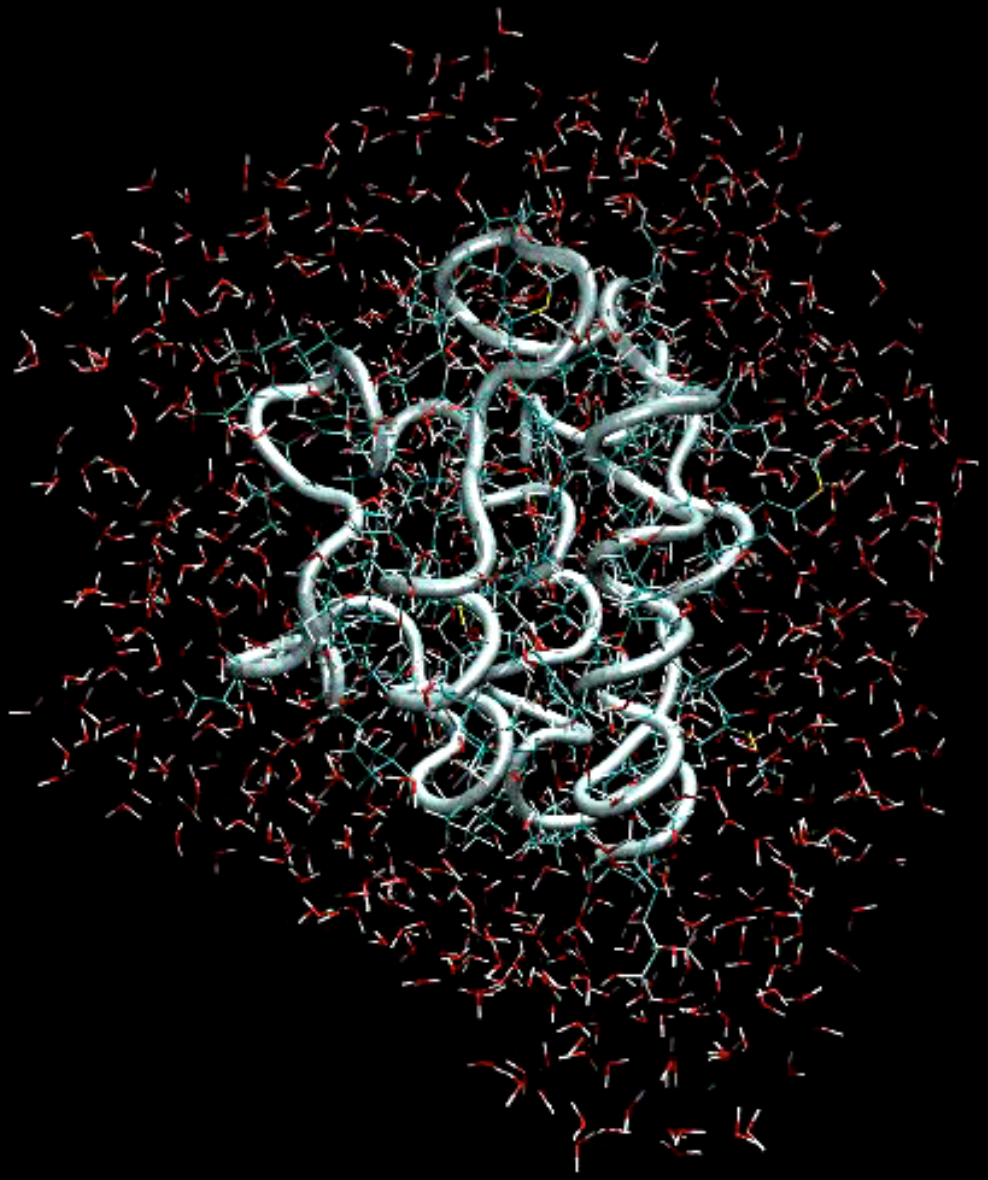
Final refinement in H2O



Automatic in ARIA

RE-COORD data base

Final refinement in H2O

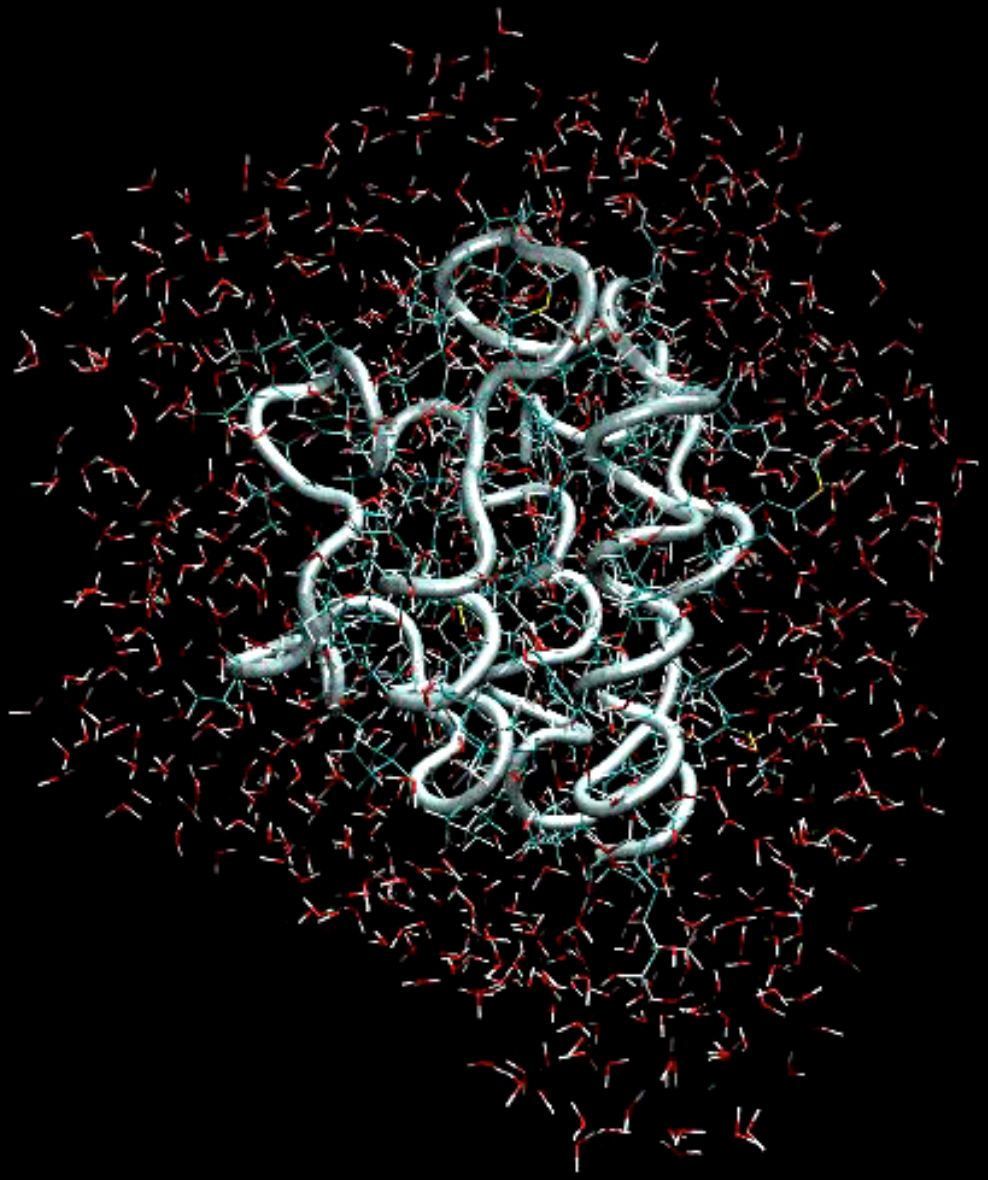


Automatic in ARIA

RE-COORD data base

Improvement of:

Final refinement in H2O



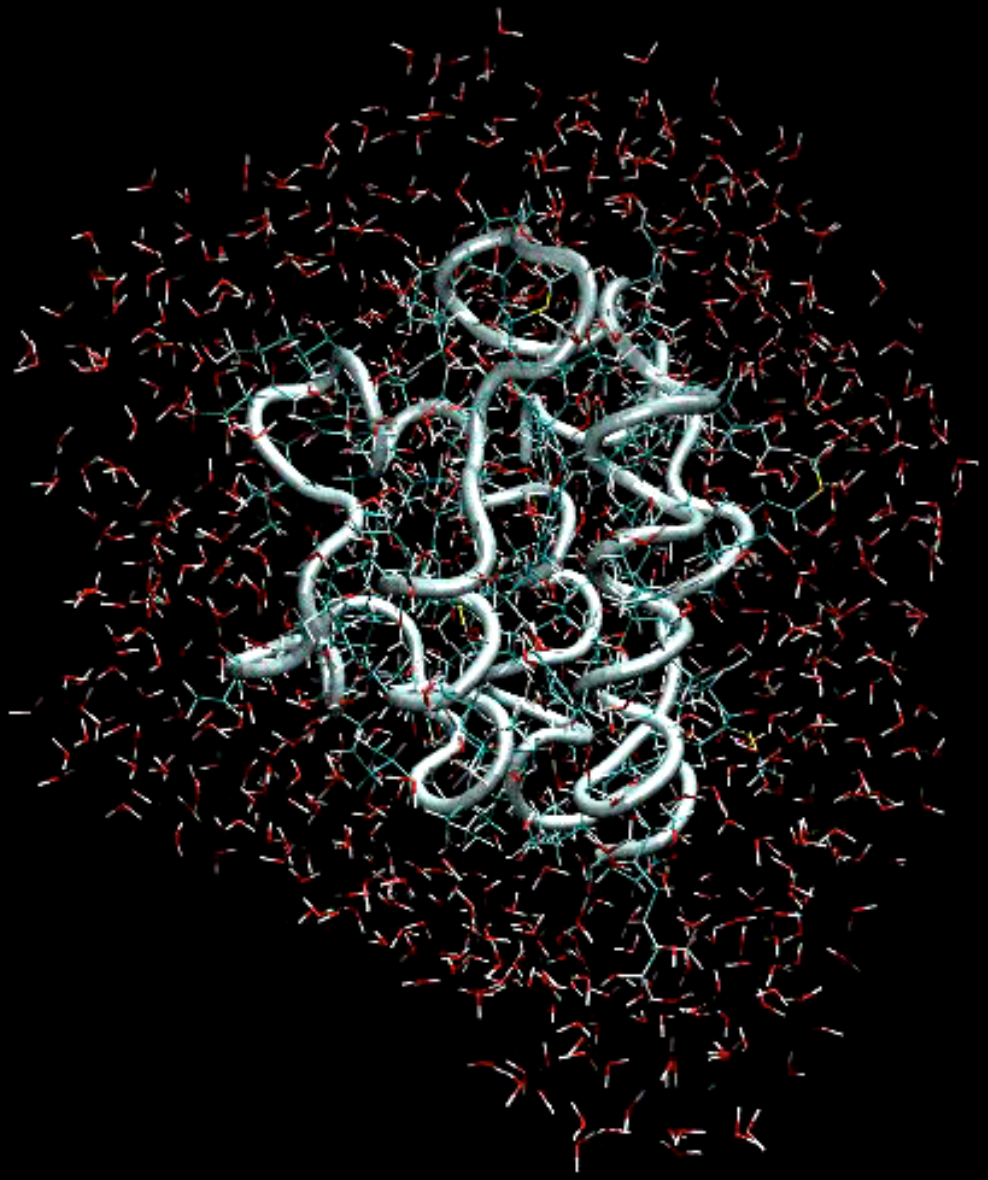
Automatic in ARIA

RE-COORD data base

Improvement of:

- validation results (phi/psi, ...)

Final refinement in H2O



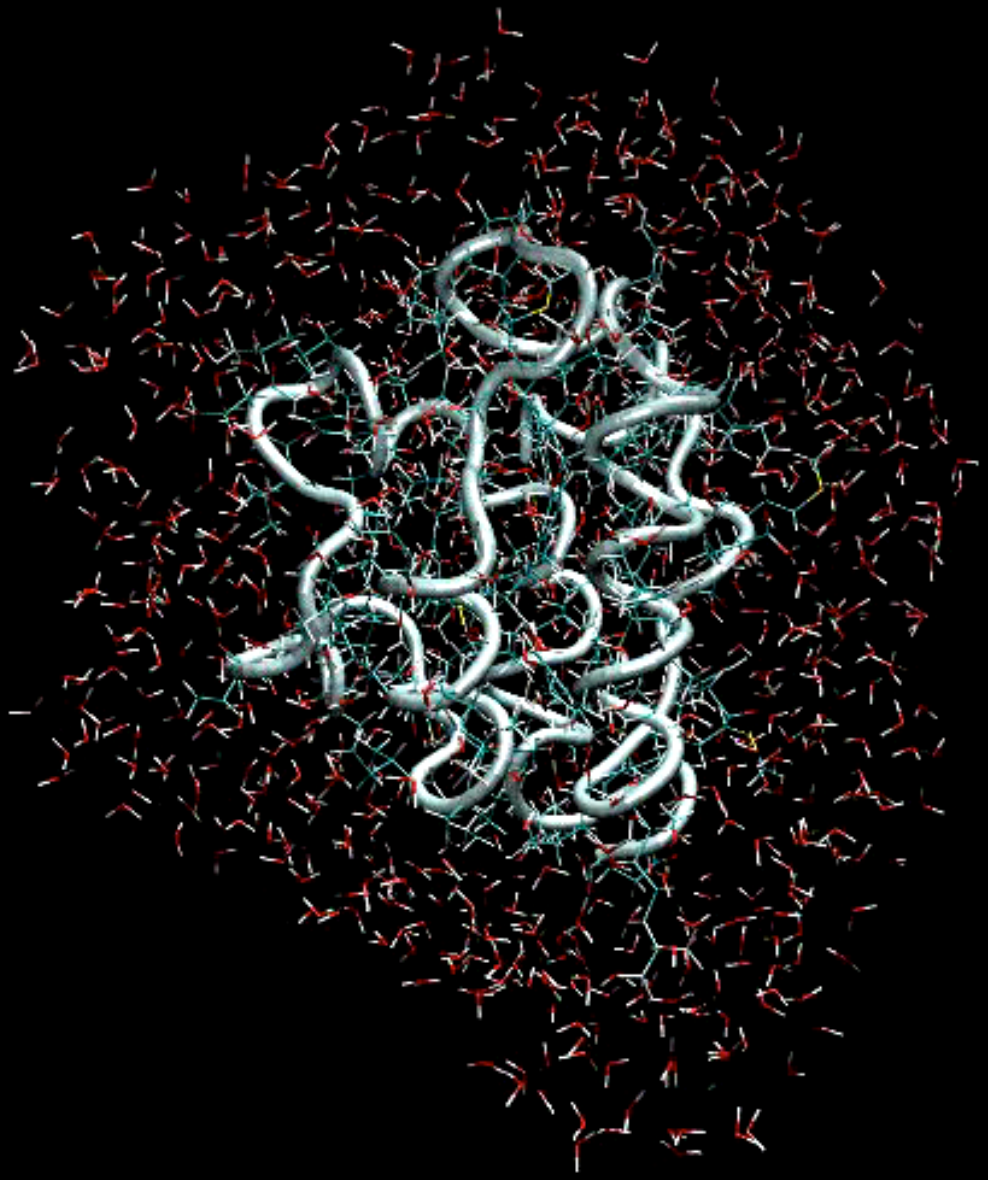
Automatic in ARIA

RE-COORD data base

Improvement of:

- validation results (phi/psi, ...)
- Q-factors

Final refinement in H2O



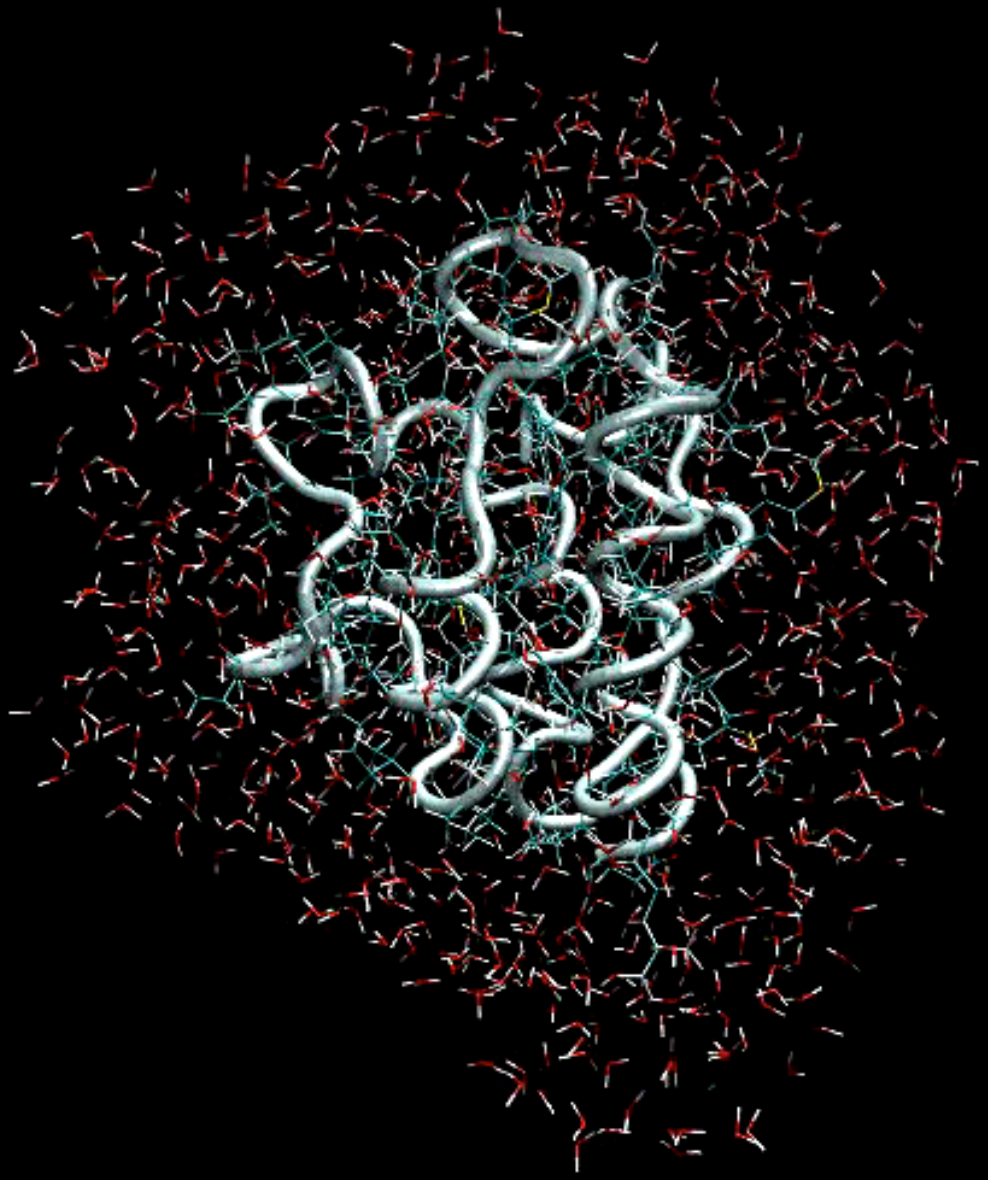
Automatic in ARIA

RE-COORD data base

Improvement of:

- validation results (phi/psi, ...)
- Q-factors
- radius of gyration

Final refinement in H2O



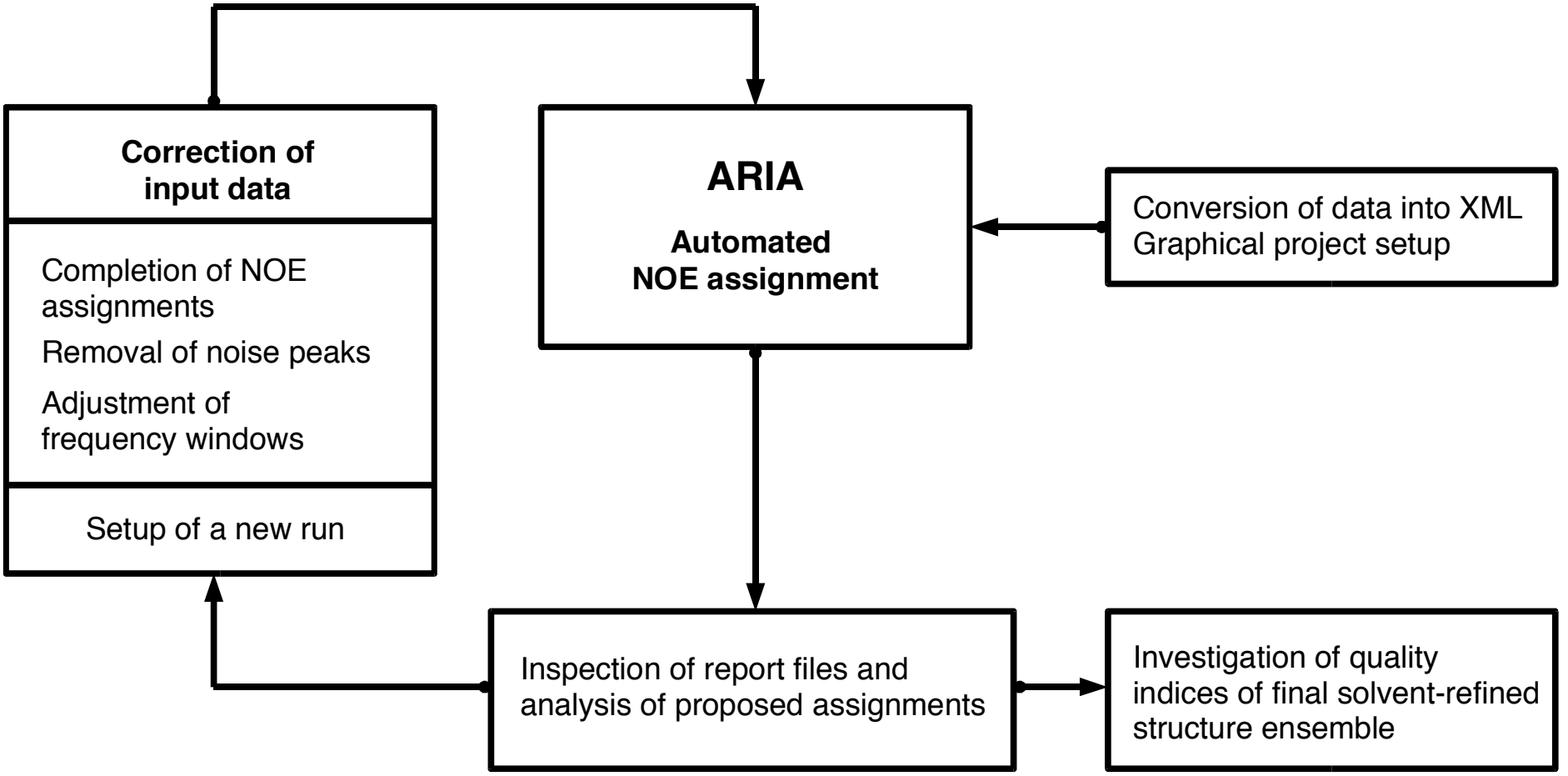
Automatic in ARIA

RE-COORD data base

Improvement of:

- validation results (phi/psi, ...)
- Q-factors
- radius of gyration
- X-ray molecular replacement

Iterations of iterations of itereations ...



Overview

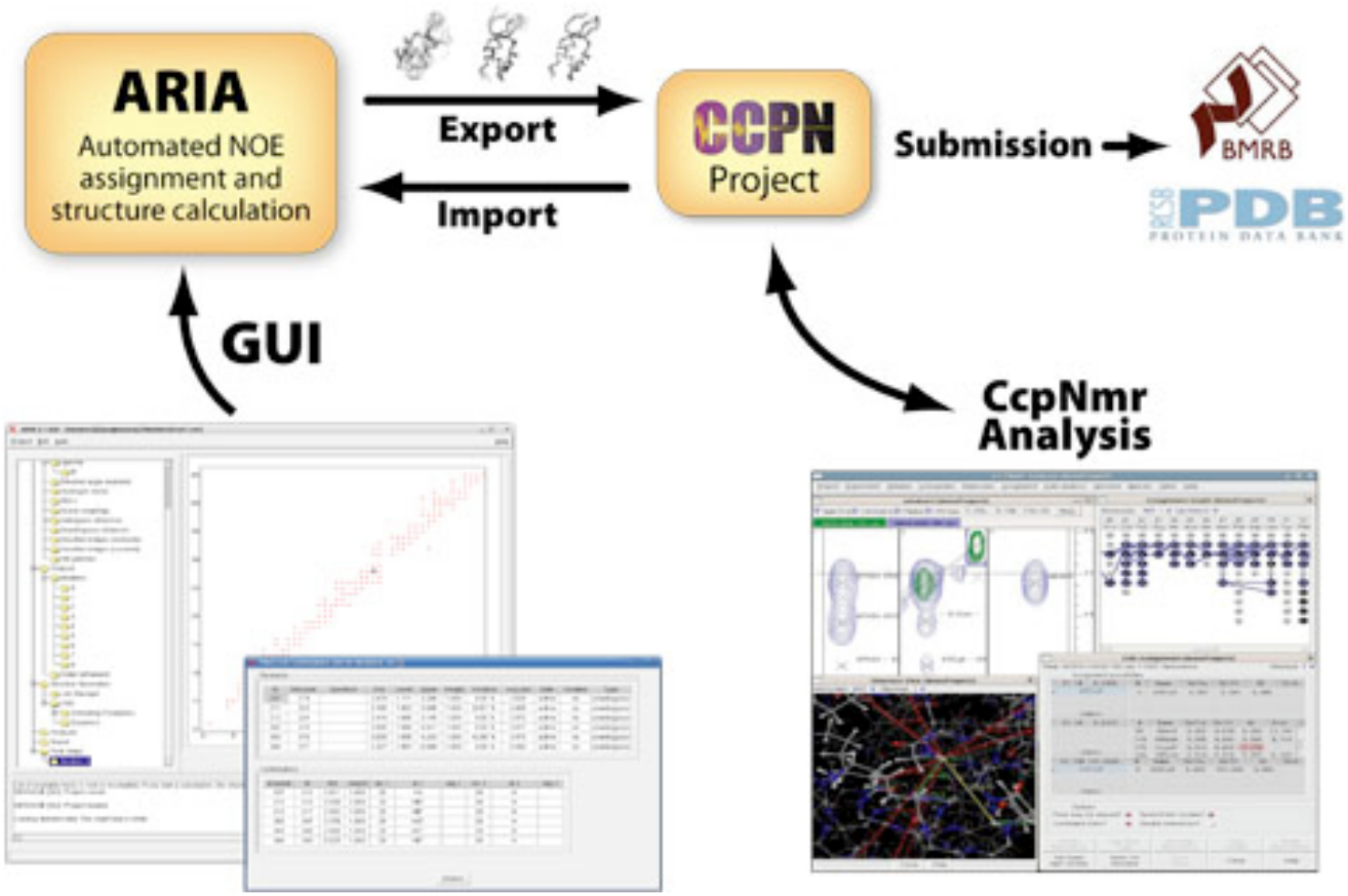
(1) Some basic concepts and history

(2) Recent developments

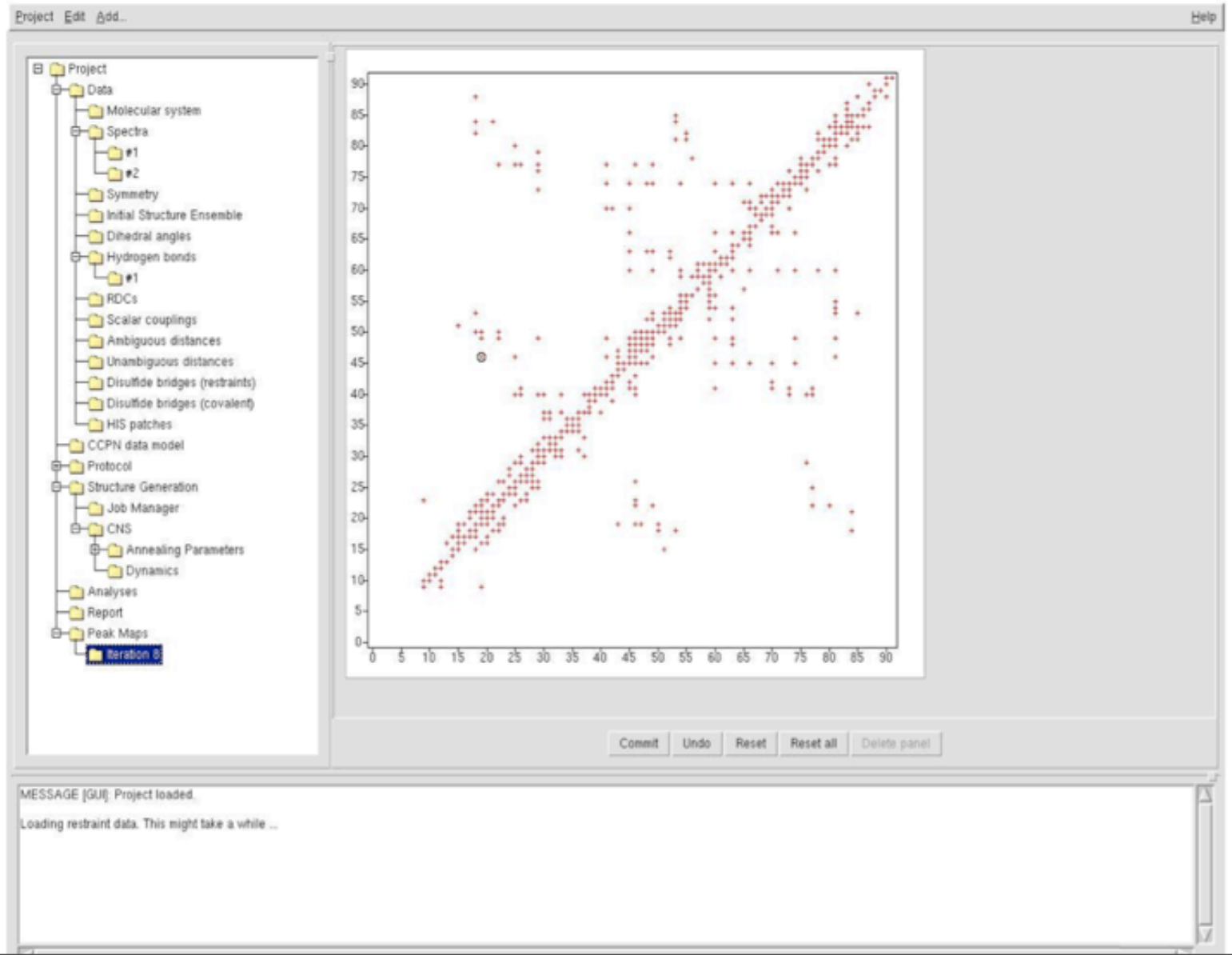
- GUI development
- Single minimum potentials
- Symmetric structures
- Solid state NMR
- More flexible structure calculation engine (CNS, YASARA, ICMD; to come: Xplor-NIH)

(3) Practical issues

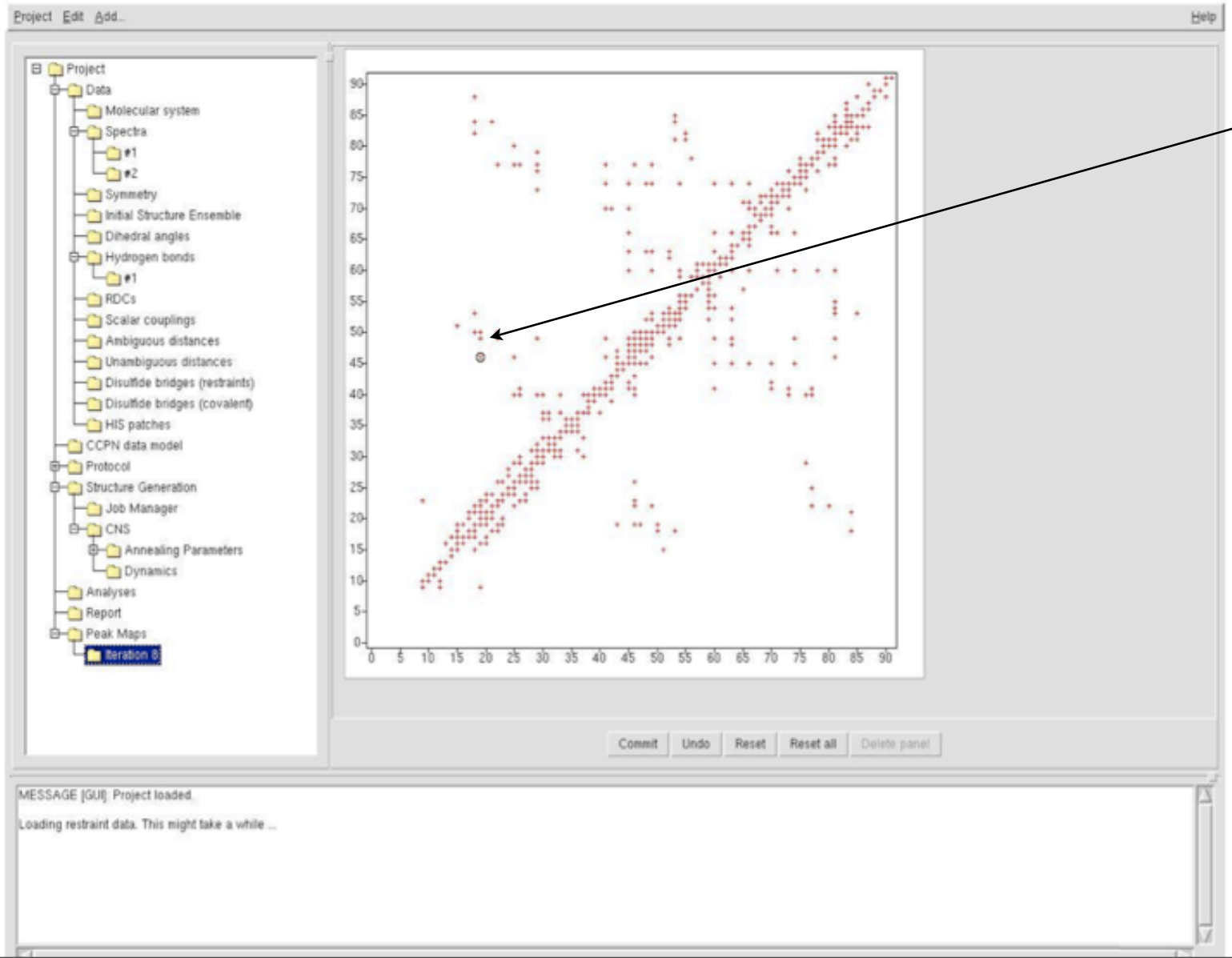
Current release: ARIA2.2



Interactive contact map



Interactive contact map



click

Assignment report for residue pair

Restraints

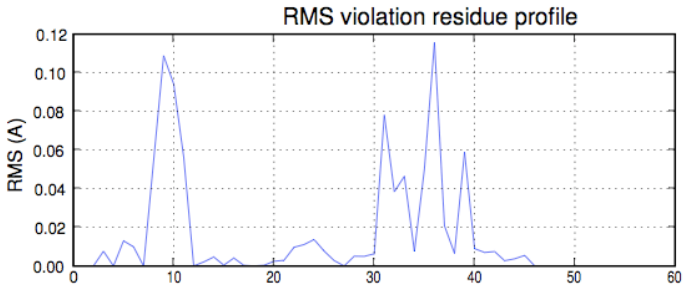
Id	Ref peak	Spectrum	Dist	Lower	Upper	Weight	Violation	Avg dist	State	Violated	Type
2327	1981	13C NOESY	3.500	1.969	5.032	1.000	0.00 %	4.134	active	no	unambiguous
2331	1985	13C NOESY	2.583	1.749	3.417	1.000	0.00 %	2.064	active	no	unambiguous
2333	1987	13C NOESY	2.725	1.797	3.654	1.000	0.00 %	2.094	active	no	unambiguous
2346	2000	13C NOESY	2.973	1.868	4.077	1.000	0.00 %	2.679	active	no	unambiguous

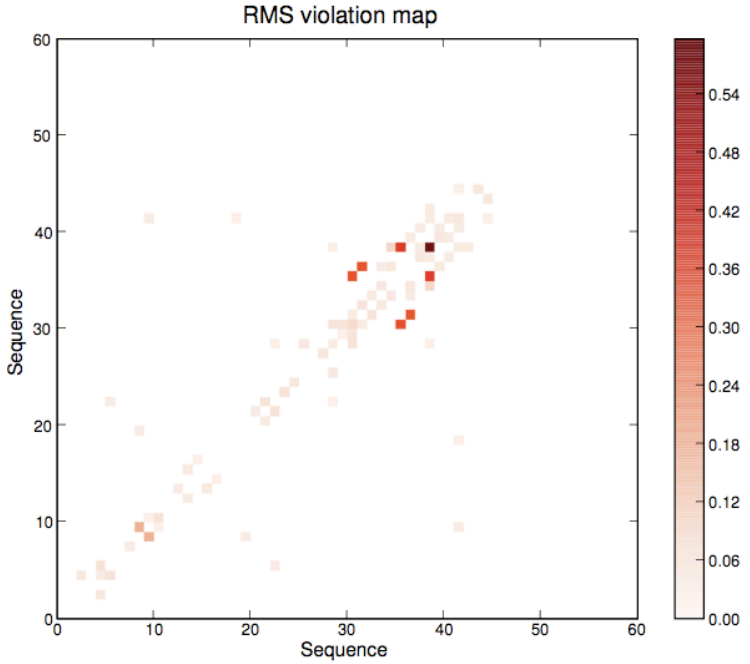
Contributions

ariapeak	id	dist	weight	res 1	at 1	seg 1	res 2	at 2	seg 2
2327	3388	4.166	1.000	19	HE*		46	HA	
2331	3392	2.180	1.000	19	HE*		46	HB*	
2333	3394	2.090	1.000	19	HE*		46	HB*	
2346	3415	2.743	1.000	19	HD*		46	HB*	

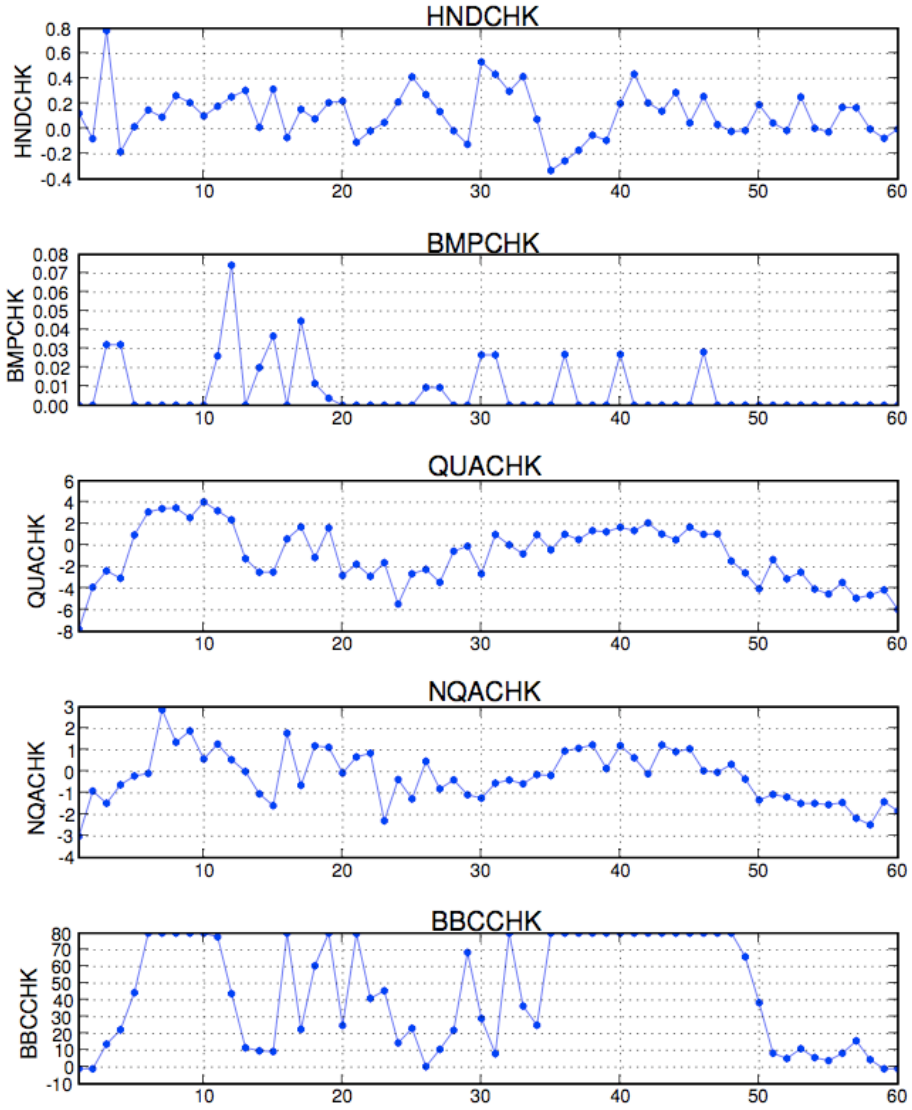
Dismiss

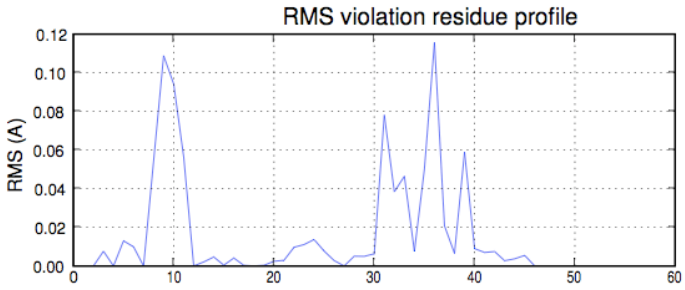
Restraint violation in ARIA

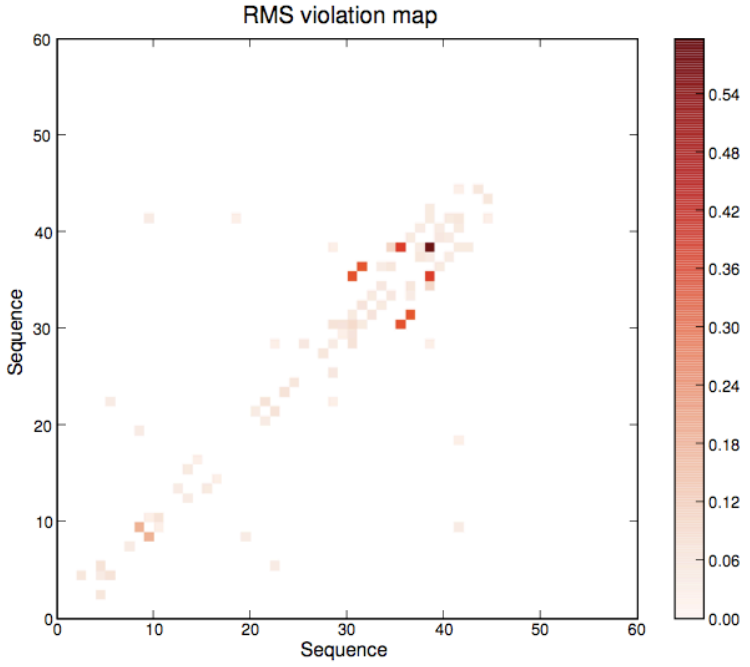




Validation results from Whatlf in ARIA







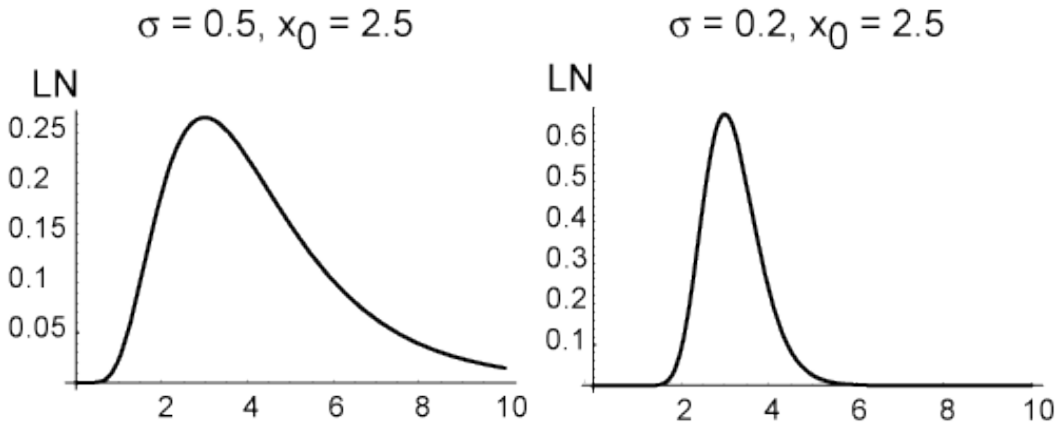
Pseudo-Bayesian structure calculation

- aim: systematically re-evaluate NMR structures with Bayesian criteria; calculation with a fast method
- data base of NMR structures with more meaningful statistics (RMSD, RMS to data)
- many problems (CPU time necessary, data conversion, ...)
- pseudo-Bayesian:
  - use Bayesian restraint potential
  - use Bayesian iterative determination of most important nuisance parameter
  - standard minimization

Bayesian restraint potential

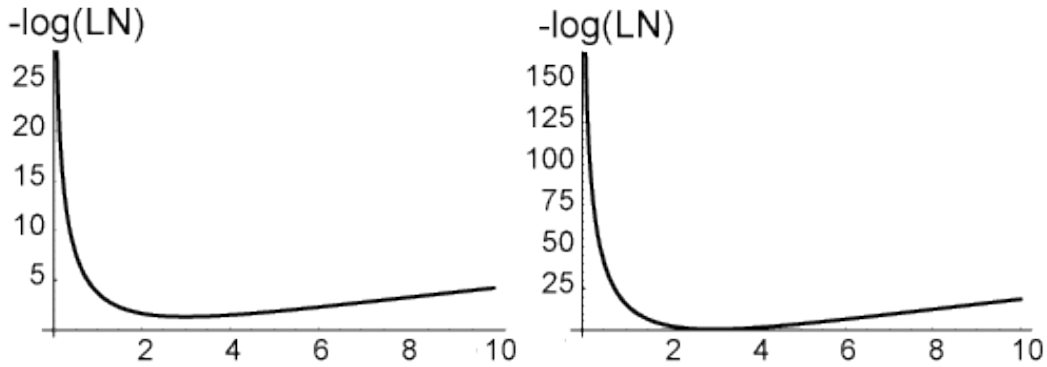
$$E = (\log[d] - \log[d_0])^2 = \left(\log\left[\frac{d}{d_0}\right]\right)^2$$

LogNormal distribution of distances



LogHarmonic restraint potential

- single minimum
- no bounds

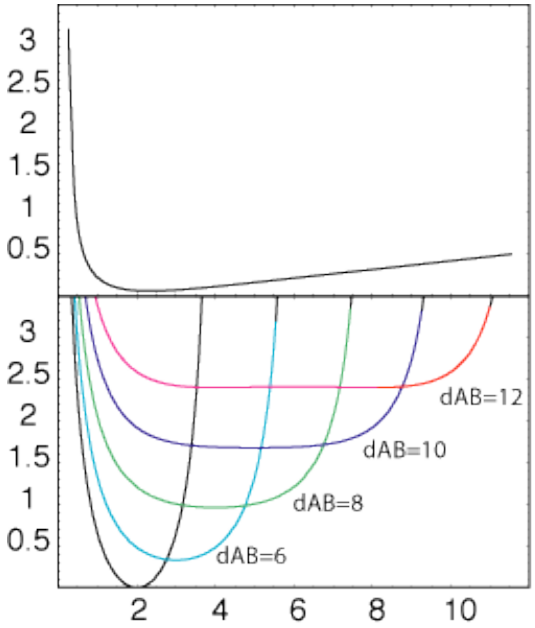


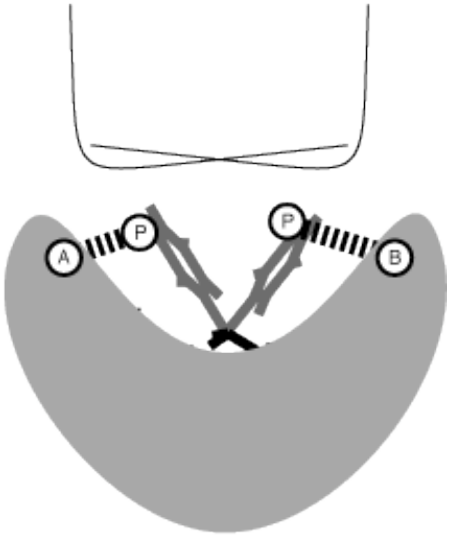
Rieping W, Habeck M, Nilges M.. J Am Chem Soc. 2005 Nov 23;127:16026-7.

Why single minimum potentials

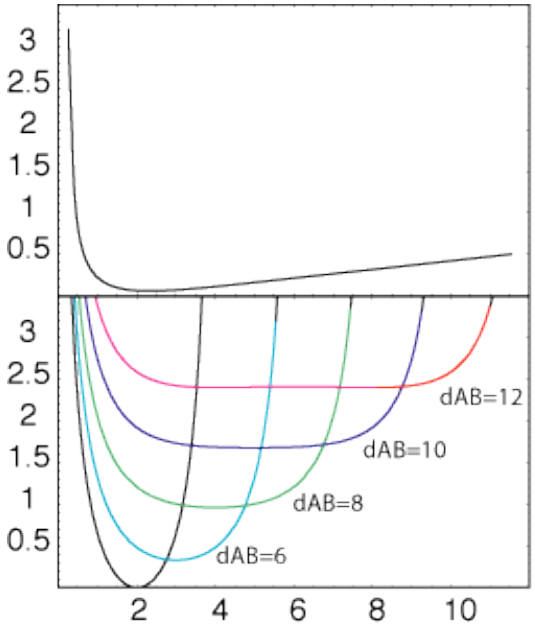
- Standard way to treat "errors": wide and generous bounds
  - conservative estimate on data quality
  - but: loss of information, introduction of additional parameters
- Alternative: single minimum potential
  - derived from Bayesian treatment (-> ISD)
  - closer to "raw data"
  - no free parameters

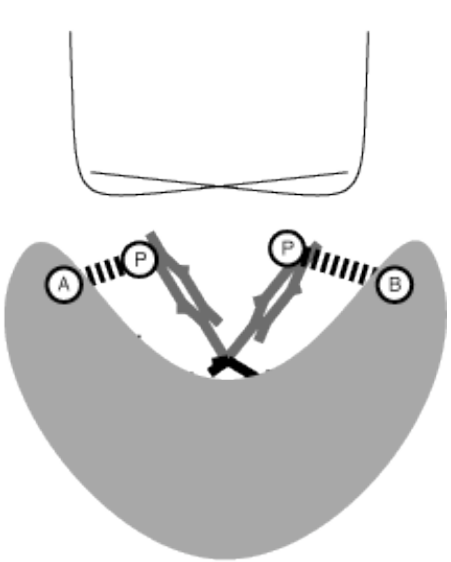
- NOEs and distances follow log-normal distribution well
- derived "log-harmonic" potential has single minimum and is very tolerant to outliers
- desired behaviour for inconsistent data
- allows for automated force constant determination

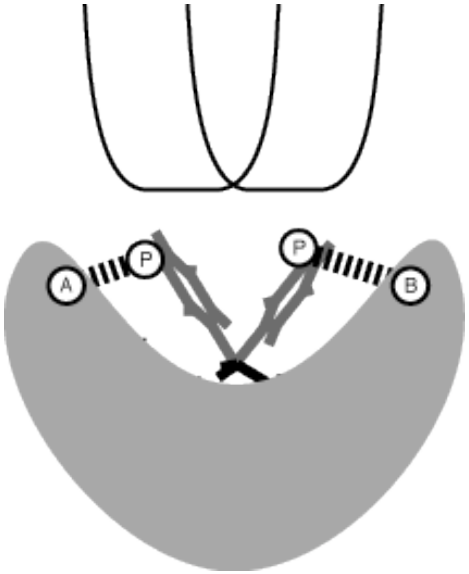




- NOEs and distances follow log-normal distribution well
- derived "log-harmonic" potential has single minimum and is very tolerant to outliers
- desired behaviour for inconsistent data
- allows for automated force constant determination







Ideal weight determined by Bayesian analysis

$$E_{hybrid} = E_{phys} + w_{data} E_{data}$$

- Bayesian analysis allows for (approximate) determination of optimal overall weight (with known structure):

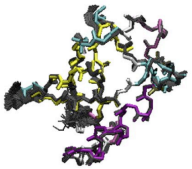
$$w_{data} = 1/RMS_{data}^2$$

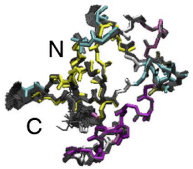
- for *determination*, update iteratively during structure calculation
- equally applicable for any least-square E_{data}
- weight ⇔ overall data quality

Habeck M, Rieping W, Nilges M (2006). **PNAS** 103:1756

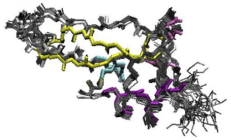
Some protein structures

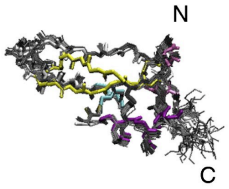
GB1



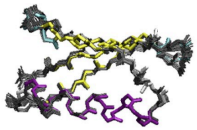


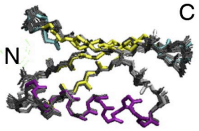
BPTI



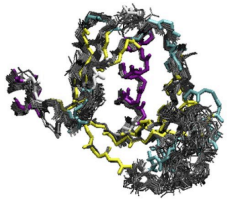


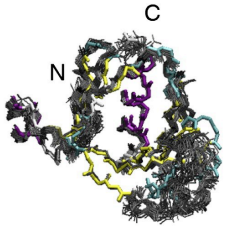
Ubiquitin

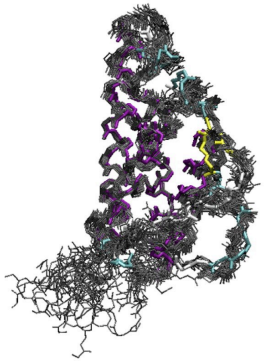


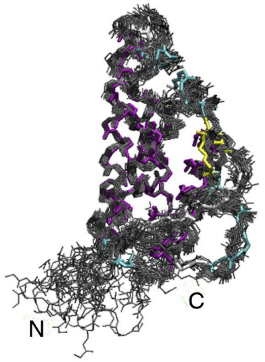


β spectrin PH domain

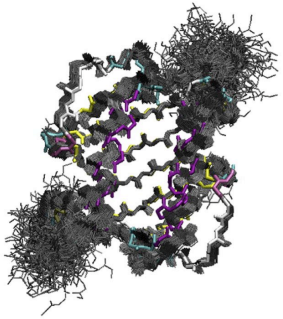


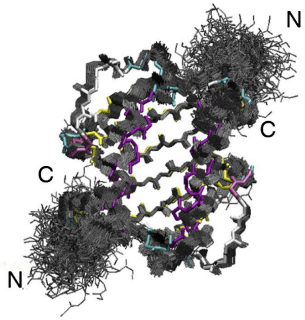






Il4



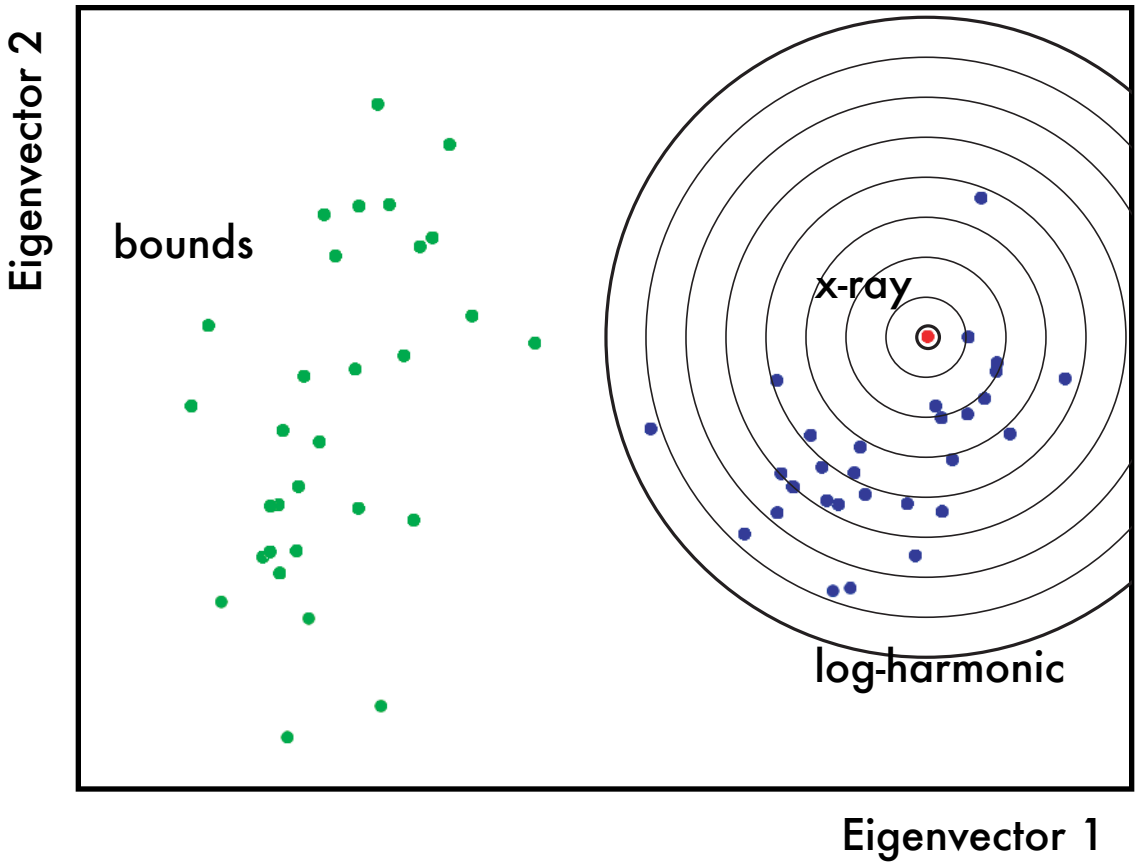


Il8

Nilges et al, Structure, 2008

Ensembles: repeated structure calculation

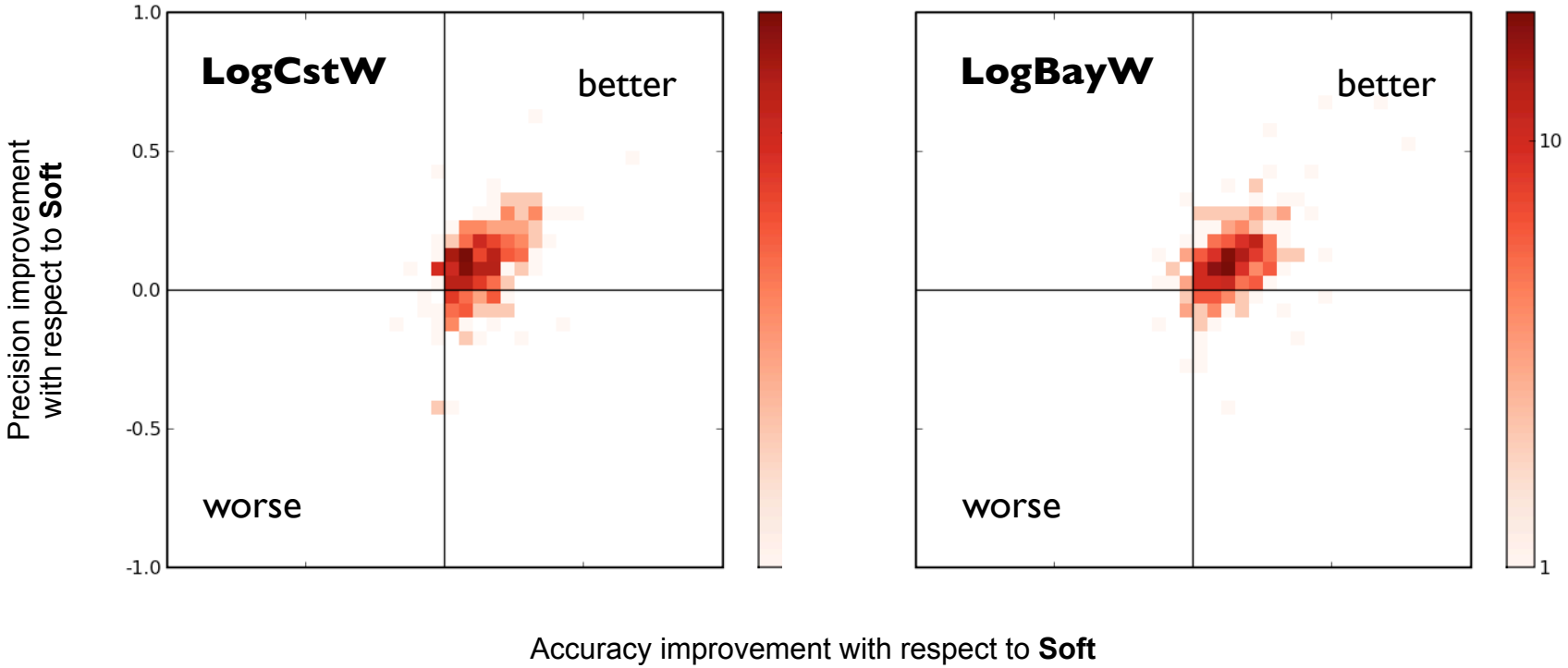
- Example IL4
  - Principle coordinate projection
  - systematically closer to X-ray
  - also: improvement of independent validation (WHATIF)



LogReCoord

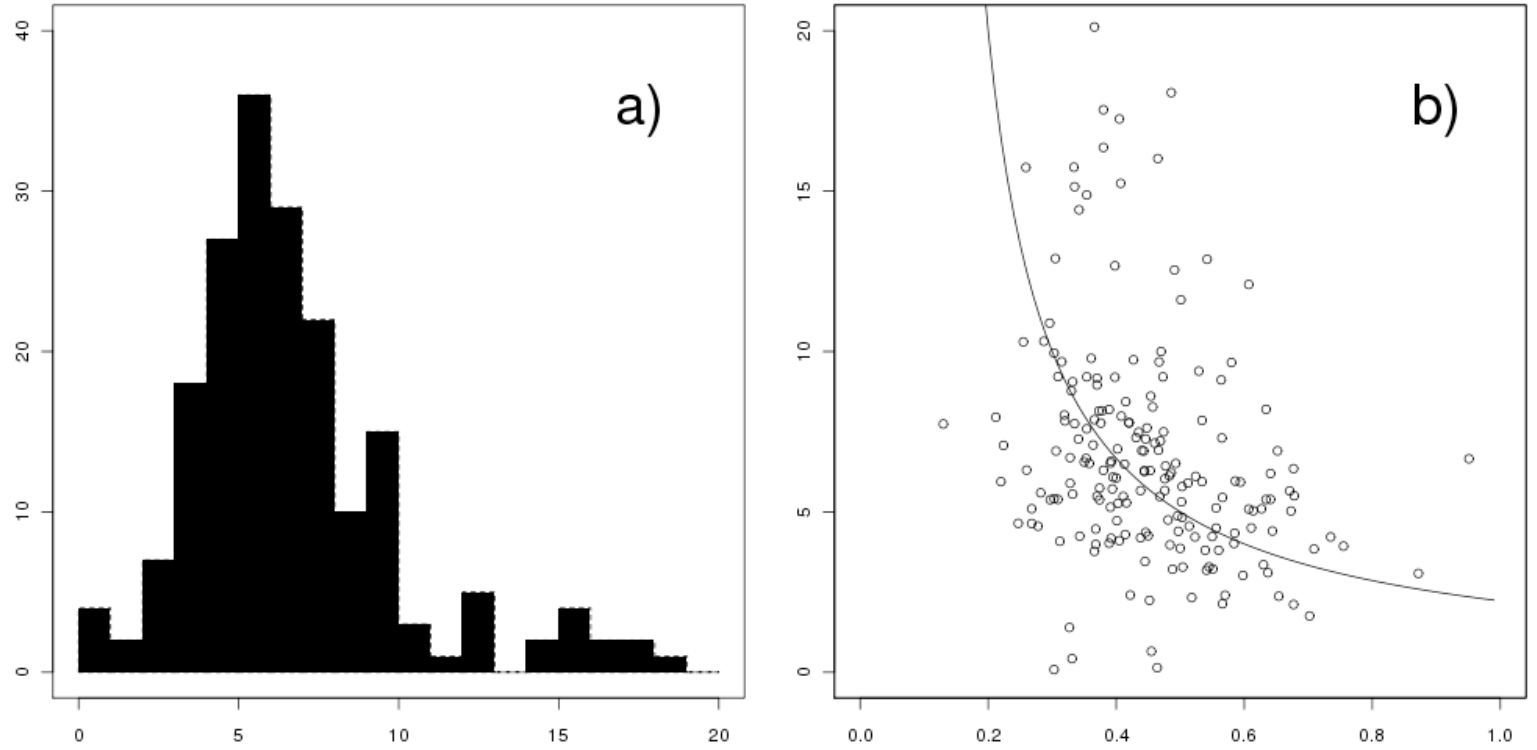
- Recalculation of 300+ NMR structures with homologous X-ray
- LogHarmonic potential with and without Bayesian weighting
- Automated handling of database restraints
  - removal of pseudo atom corrections
  - removal of other undocumented corrections
  - re-calibration to get a single target distance

Accuracy/precision



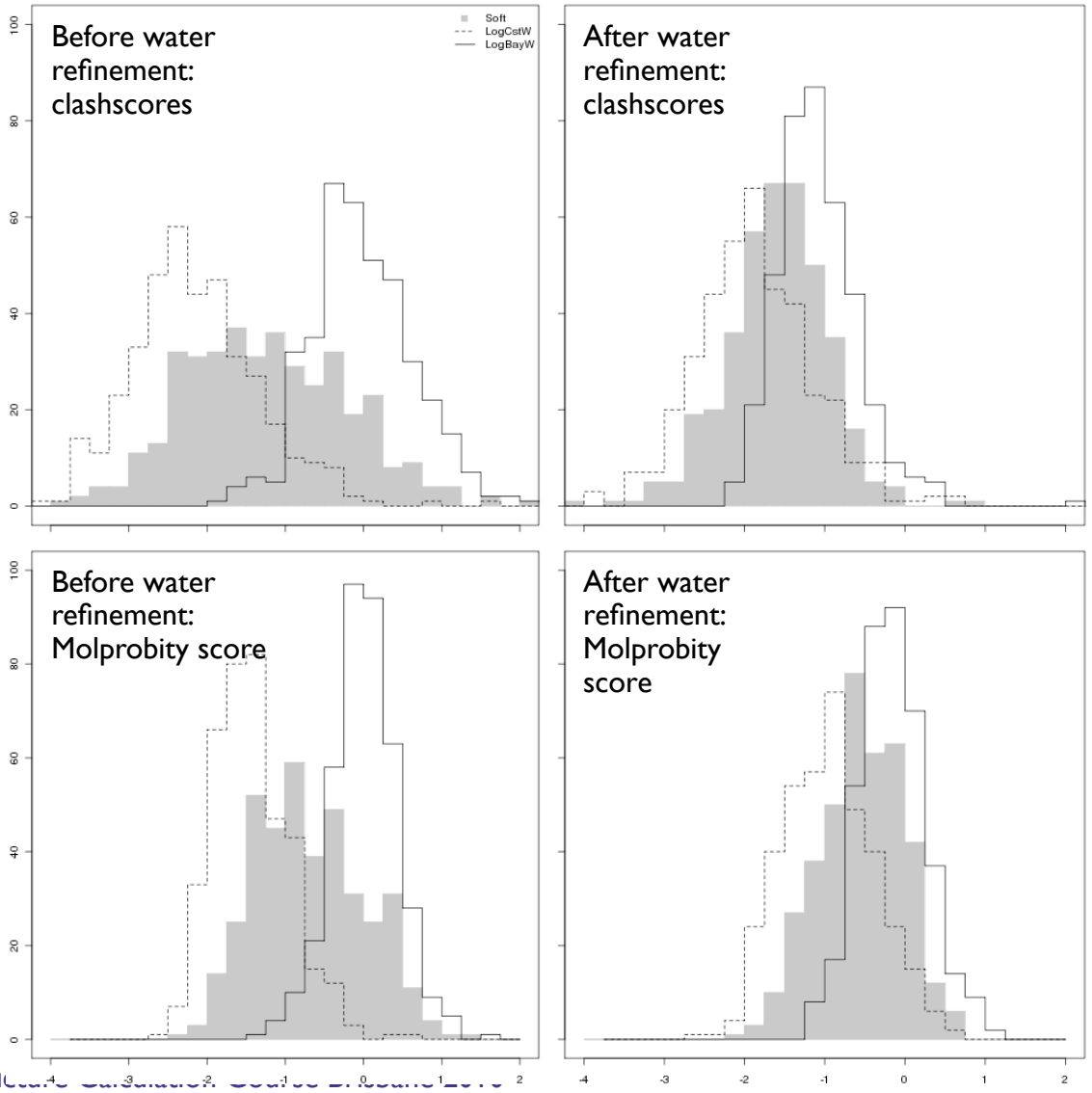
- generally, smaller RMSD and closer to X-ray

Bayesian weights



- Small weights (6 kcal/mol) for most structures
- Weak relationship between weight and accuracy

Structure quality: Molprobity



Conclusions for 300+ structures

- Successful test of LogHarmonic potential on large set of structures
  - LogRecoord database will be provided (Wim Vranken)
- Removal of bias, no modifications of bounds necessary
- Rather tight distribution of optimal weights (around 6) with some outliers (e.g., Ubiquitin, GB1)
- Improvements of structures:
  - Major and systematic increase in precision and similarity to X-ray structure
  - Better fit to restraints (if logarithmic metric is used), but more outliers
  - Improvement of WhatIF and MolProbity analysis for LogBayW calculation
- Improvements due to water refinement and LogBayW not cumulative
  - probably, low weights on NOE distances too low for current water refinement force field/ protocol

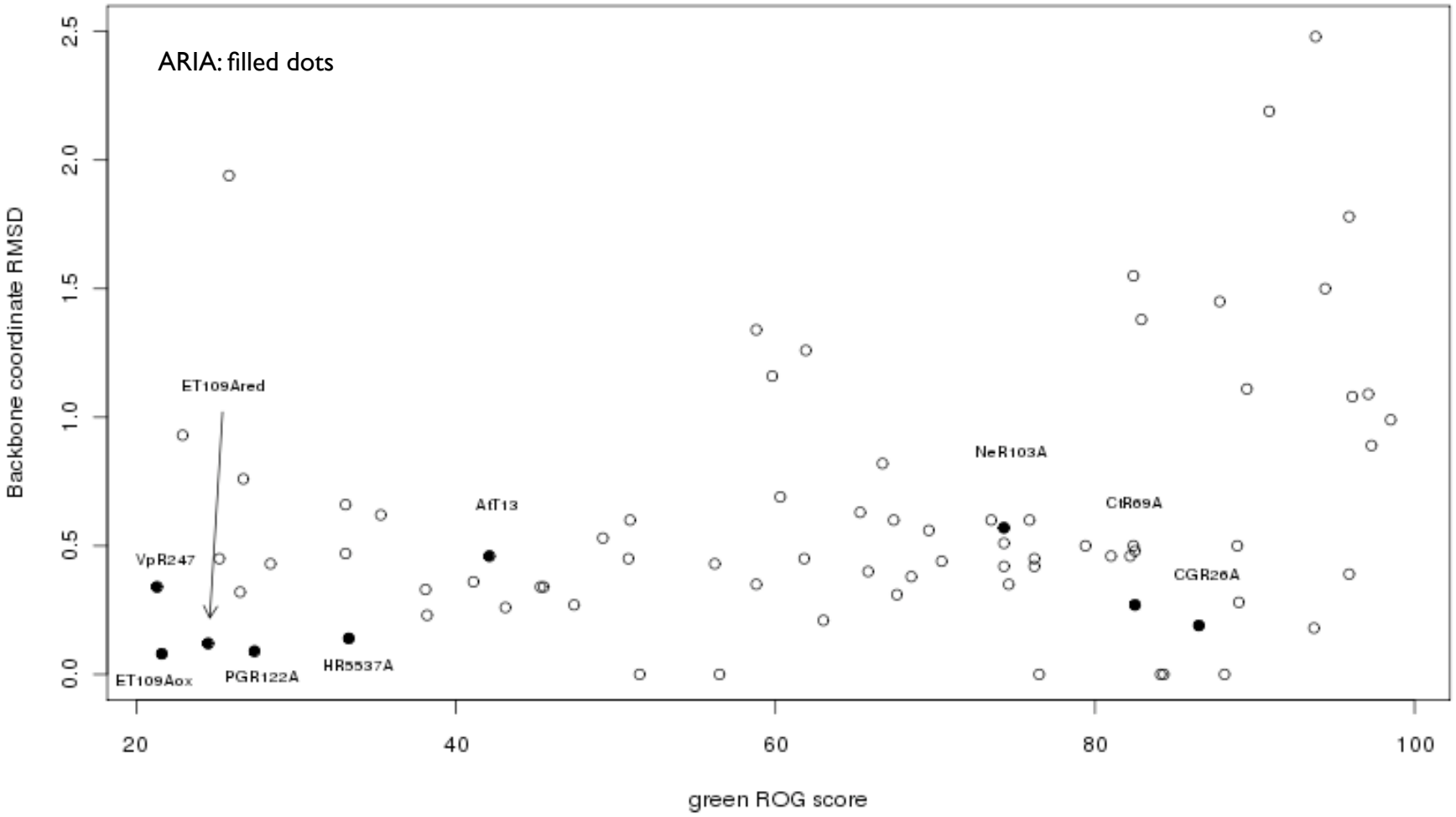
The CASD "competition" initiated by CERM

- CASD-NMR
  - Critical Assessment of Automated Structure Determination of Proteins by NMR
  - comparison of automated NMR protein structure calculation methods
  - CASD-NMR open for any laboratory to participate and/or to submit targets
  - community-wide rolling experiment
  - CASD-NMR participants have eight weeks to generate structures
- First round
  - 10 datasets, most of them with known structures (not blind for most)
  - first meeting in Florence, Italy on May 4-6, 2009
- Second round
  - 8 new datasets, blind testing on targets provided by several groups
  - second meeting in Florence, Italy on May 5-7, 2010
- Validation by Geerten Vuister / Jurgen Doreleijers

LogHarmonic potential in ARIA and CASD

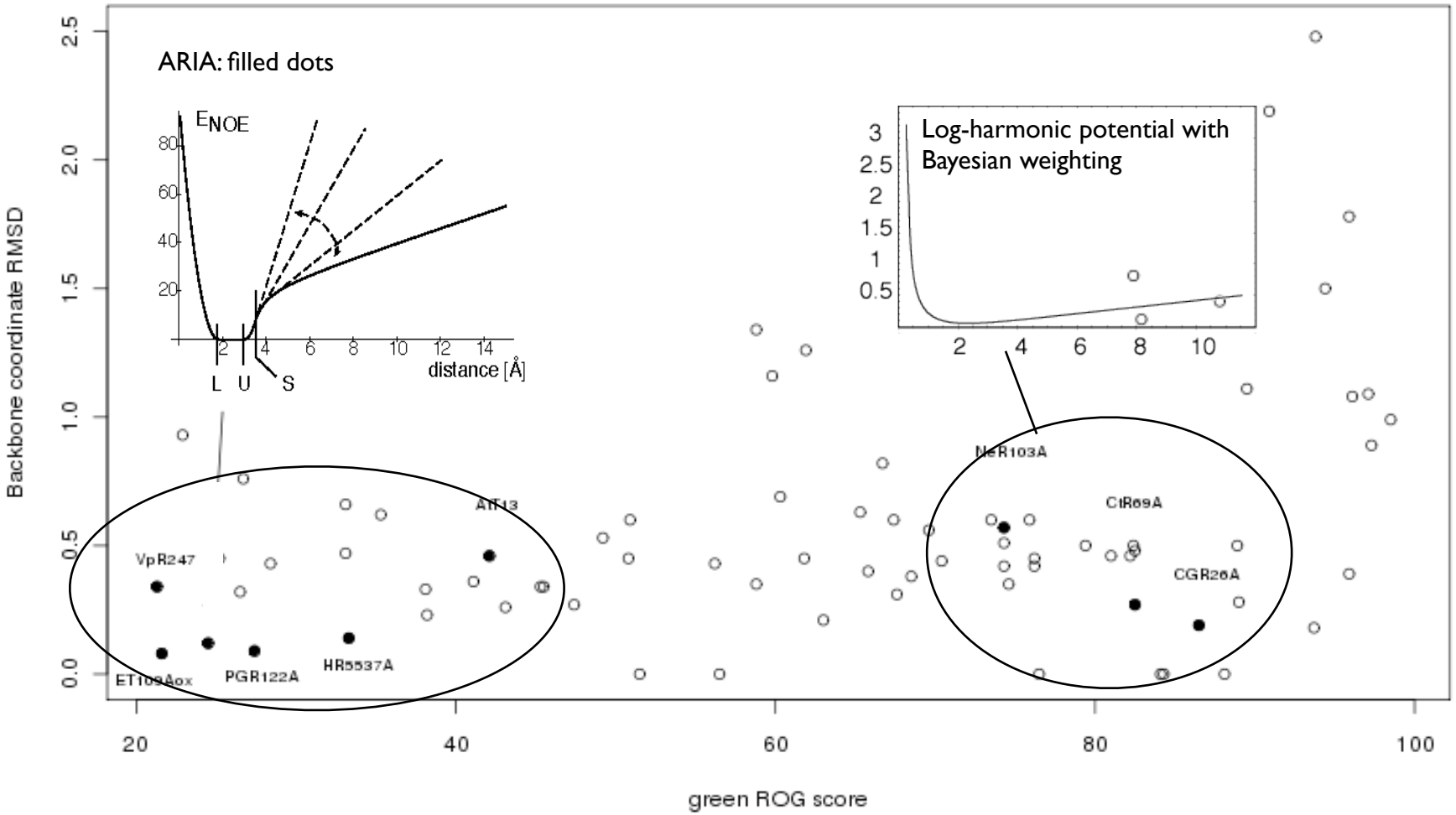
- Different task:
  - much more ambiguities in the data
  - potential false restraints, exclusion of restraints possible
- Implementation:
  - only in cooling phase, as previously
  - rules for assignment of restraints not changed
  - rules for exclusion of restraints not changed (i.e., deviation from **bounds** estimated by ARIA)

Quality/precision for CASD targets (blind)



- Overview over all calculations, some with, some without LogBayW

Quality/precision for CASD targets (blind)

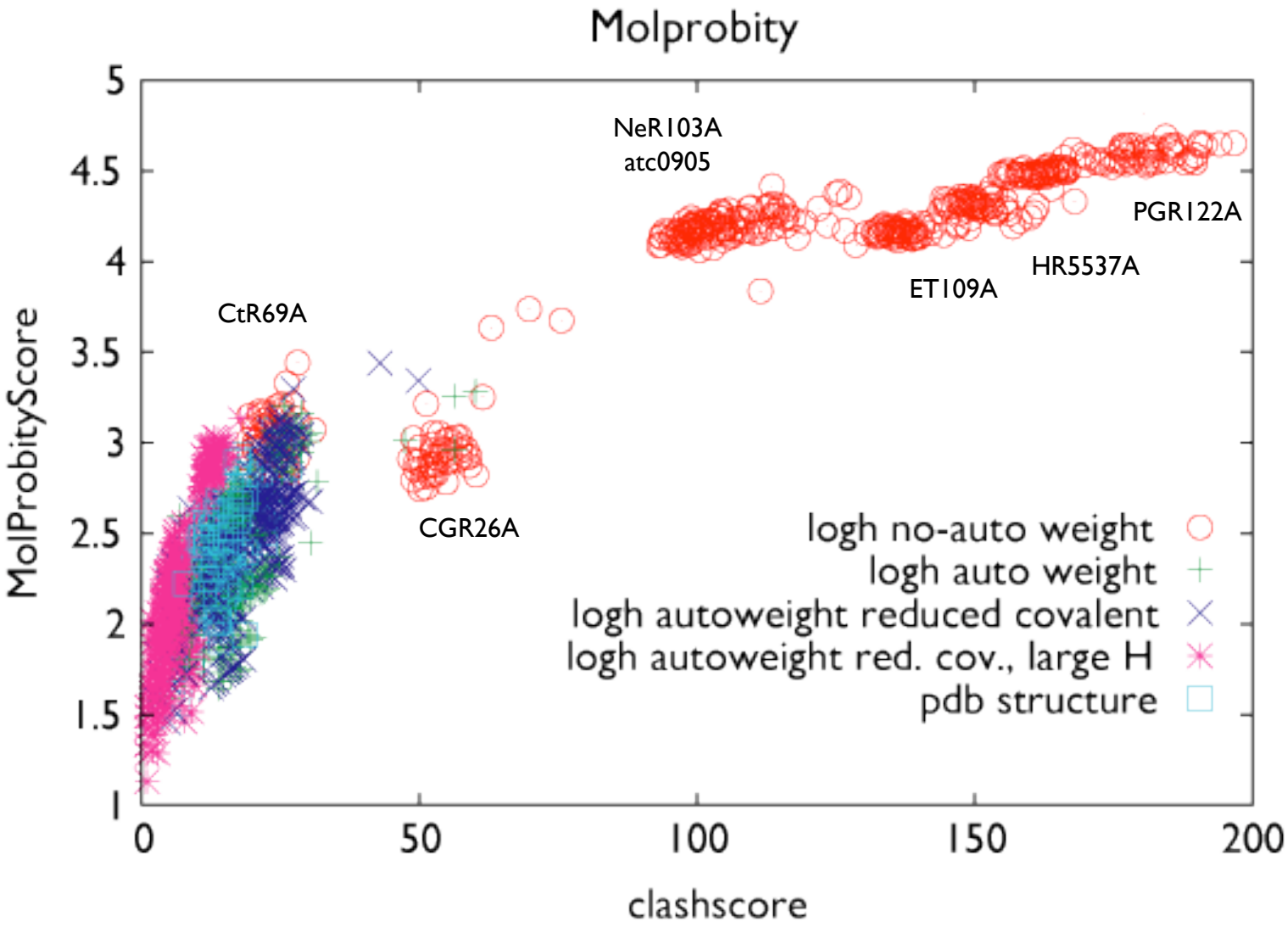


- Overview over all calculations, some with, some without LogBayW

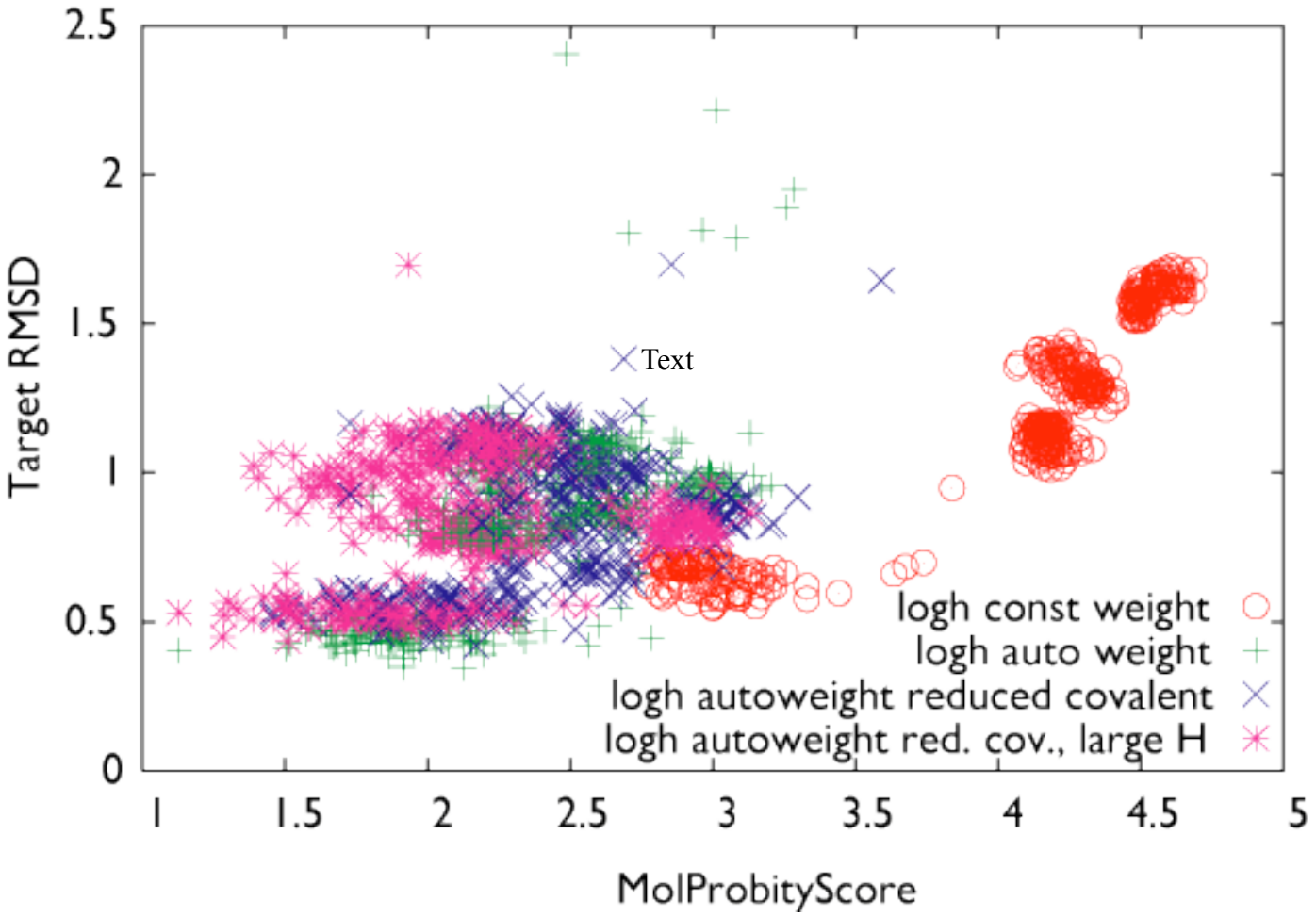
Recalculation of all CASD targets

- LogHarmonic potential and constant weight (LogCstW)
- LogHarmonic potential with Bayesian weight (LogBayW)
- LogBayW w/ reduced angle and planarity constants
  - 50 kcal/(mol rad^2) vs 500 kcal/(mol rad^2)
- LogBayW w/ reduced angle and planarity constants and big H
  - 50 kcal/(mol rad^2) vs 500 kcal/(mol rad^2)
  - H radii increased by approx. 10 %

All MolProbity Scores



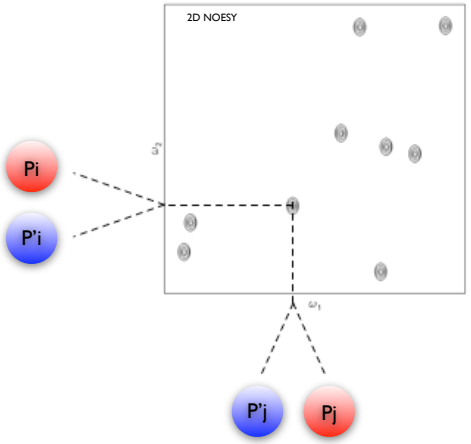
RMSD to PDB vs MolProbity



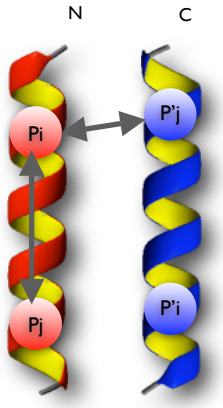
Conclusions CASD

- LogBayW significantly improves quality of ARIA structures
  - RMSD to target, WhatIf, MolProbity
  - LogBayW will be the default in ARIA
  - softer force field: RMS Z-scores
- Current / future work:
  - improve criteria for restraint removal
  - optimize number of steps / structures per iteration
  - further work on force field
  - further work on water refinement
- Implementation on GRID available

Calculation support for symmetric structures



- Increased ambiguity (identical chemical environment for all monomers)
- Experimental solution (filtered experiments) sometimes do not work



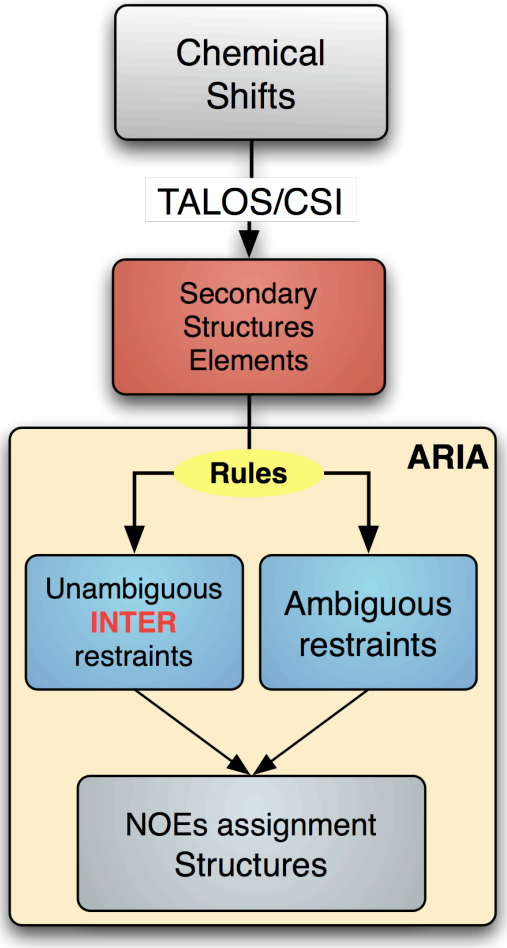
Grid search methods (with known structure)

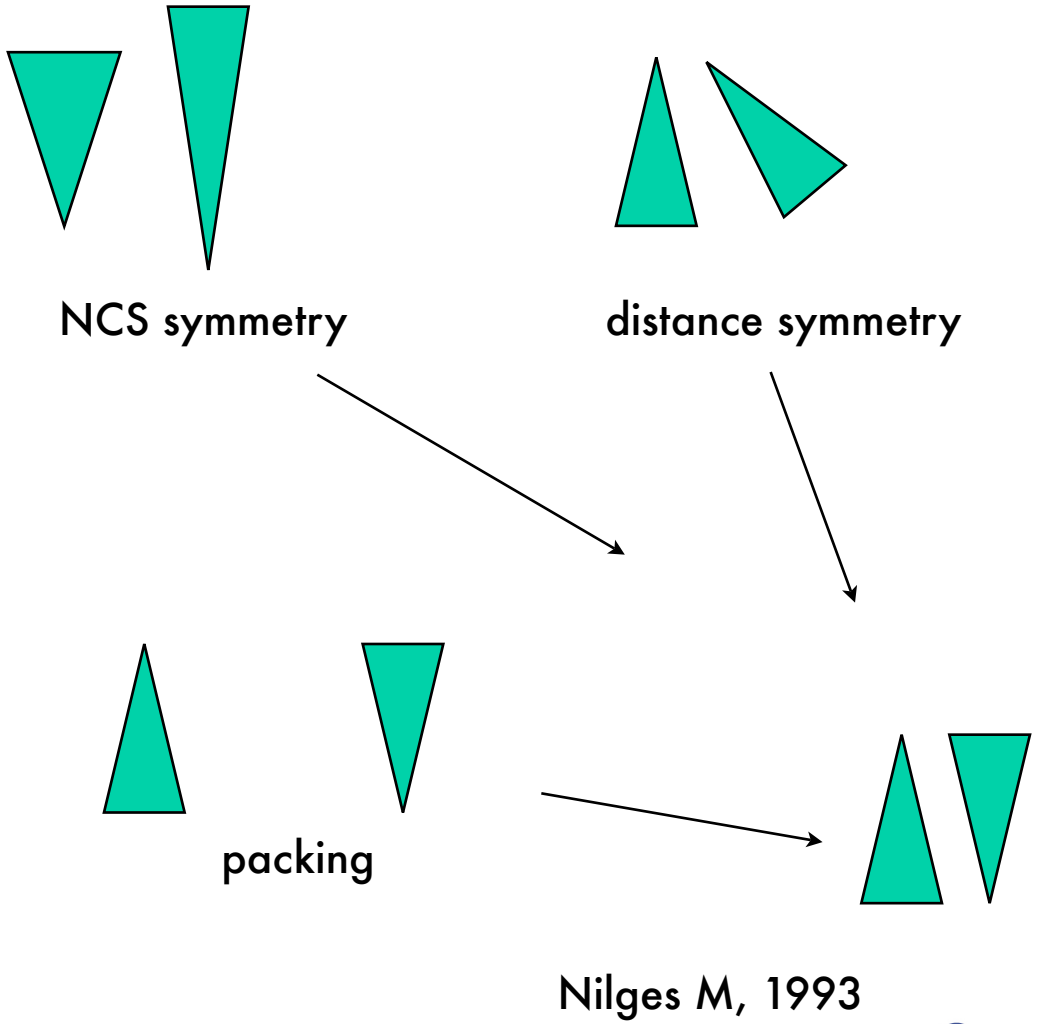
Wang CS, Lozano-Perez T, Tidor B. AmbiPack: a systematic algorithm for packing of macromolecular structures with ambiguous distance constraints. Proteins. 1998 Jul 1;32(1):26-42.

Potluri S, Yan AK, Chou JJ, Donald BR, Bailey-Kellogg C. Structure determination of symmetric homo-oligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing. Proteins. 2006 Oct 1;65(1):203-19.

- Limitations:
  - Monomer structure has to be known and rigid
  - Not general in terms of symmetry

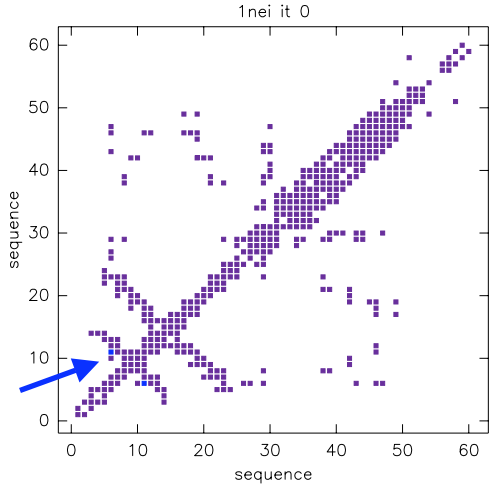
ARIA2.2: pre-filtering + symmetry restraints



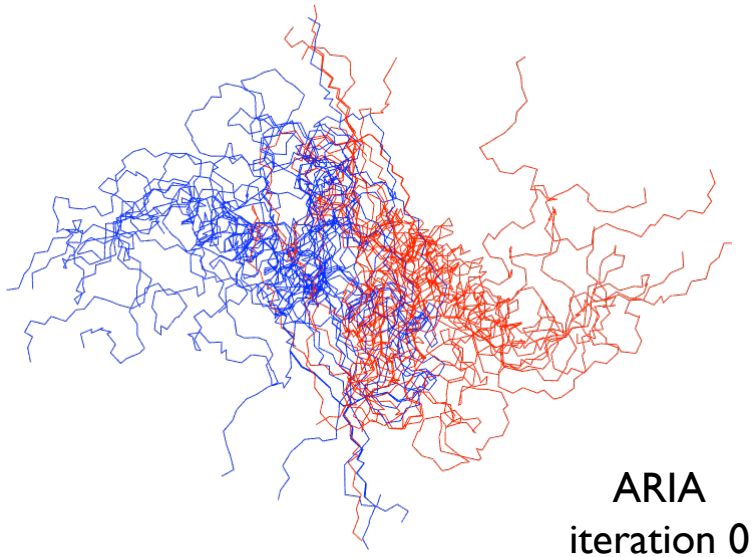


Example: Hypothetical protein Yoag

- Data from Wu et al., 2002, pdb: 1NEI
- For test, inter/ intra monomer assignments discarded
- Prefiltering: 2 unambiguous inter-residue restraints

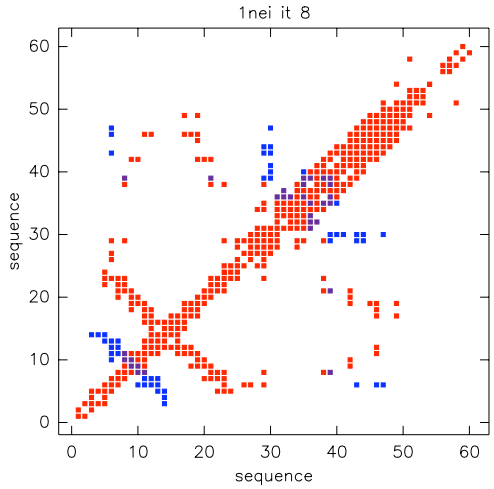


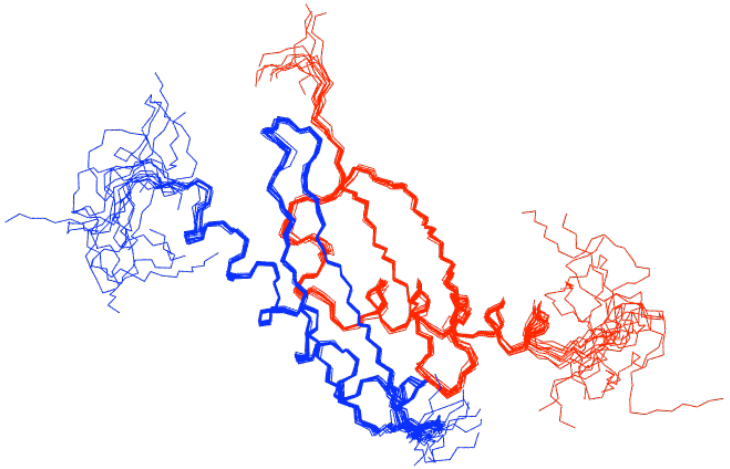
Intra/inter
Inter



ARIA
iteration 0

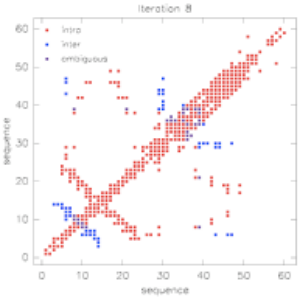
- ARIA 2.2 last iteration



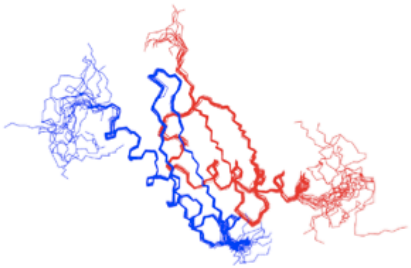


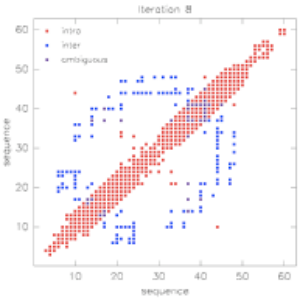
Intra/inter
Inter
Intra

more examples

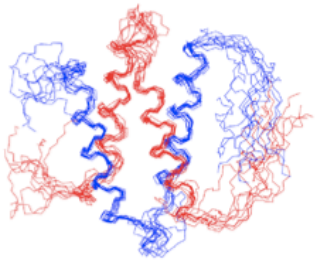


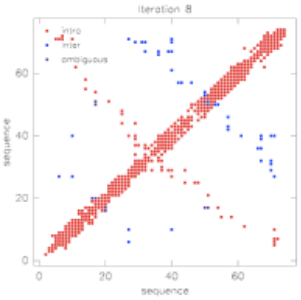
a)



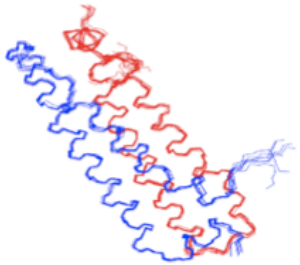


b)

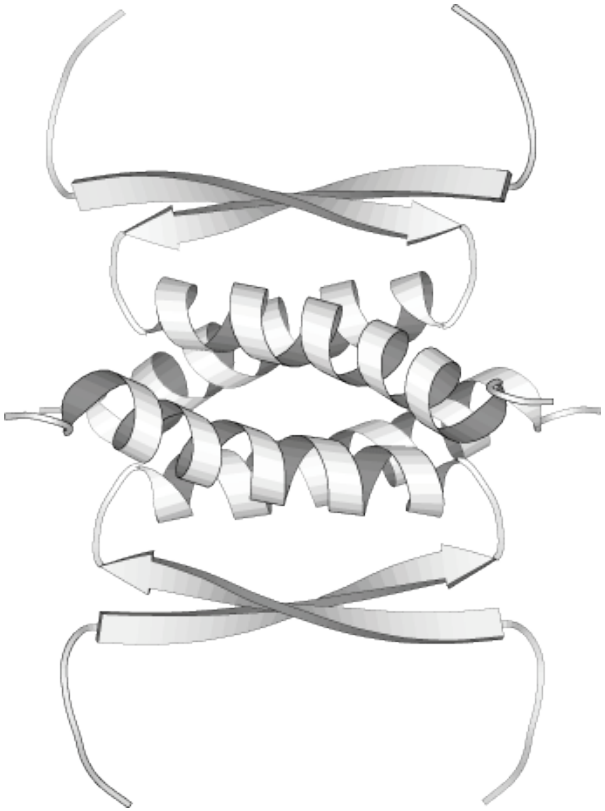




c)

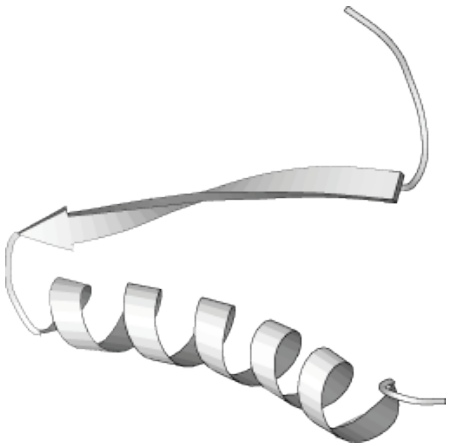


New developments: strict NCS symmetry



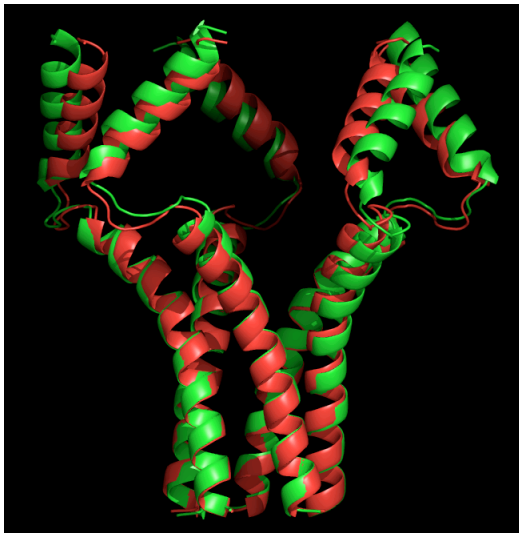
- Distance symmetry concept is limited
  - fiber symmetries only approximately
  - x-tal symmetries unfeasible (too many neighbours)
- Strict symmetry using symmetry operators

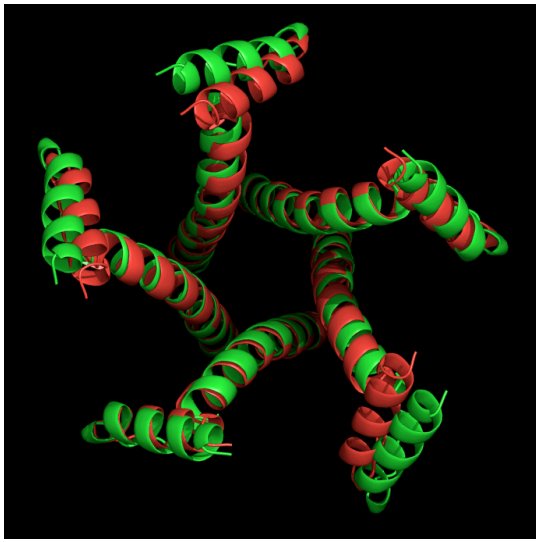
strict NCS symmetry



- calculate only monomer
- treat NOE and non-bonded with symmetry operatores
- any symmetry possible (also translational, not only C_n)
- successful test cases:
  - dynein light chain dimer
  - p53 tetramer
  - Il8 dimer
  - GB1 tetramer

Phospholamban pentamer

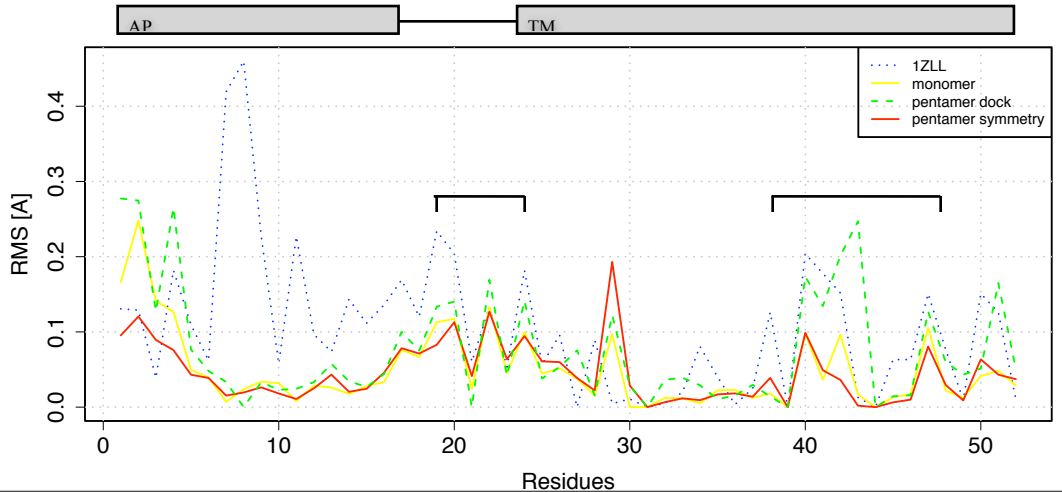




RMSD (heavy atoms)
TM domain: 1.1 A
Full pentamer: 2.2 A

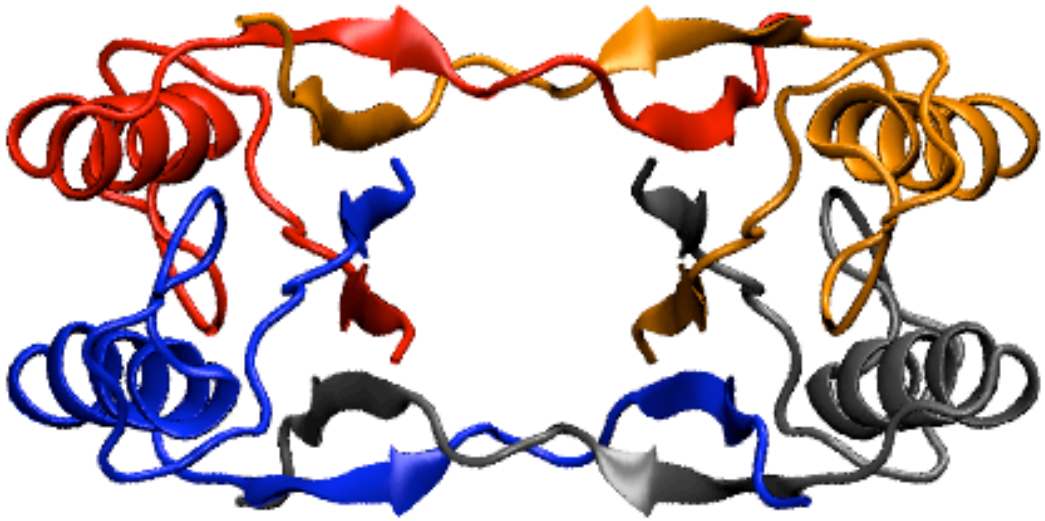
- ambiguous NOEs + RDCs

Docking method
Symmetry method



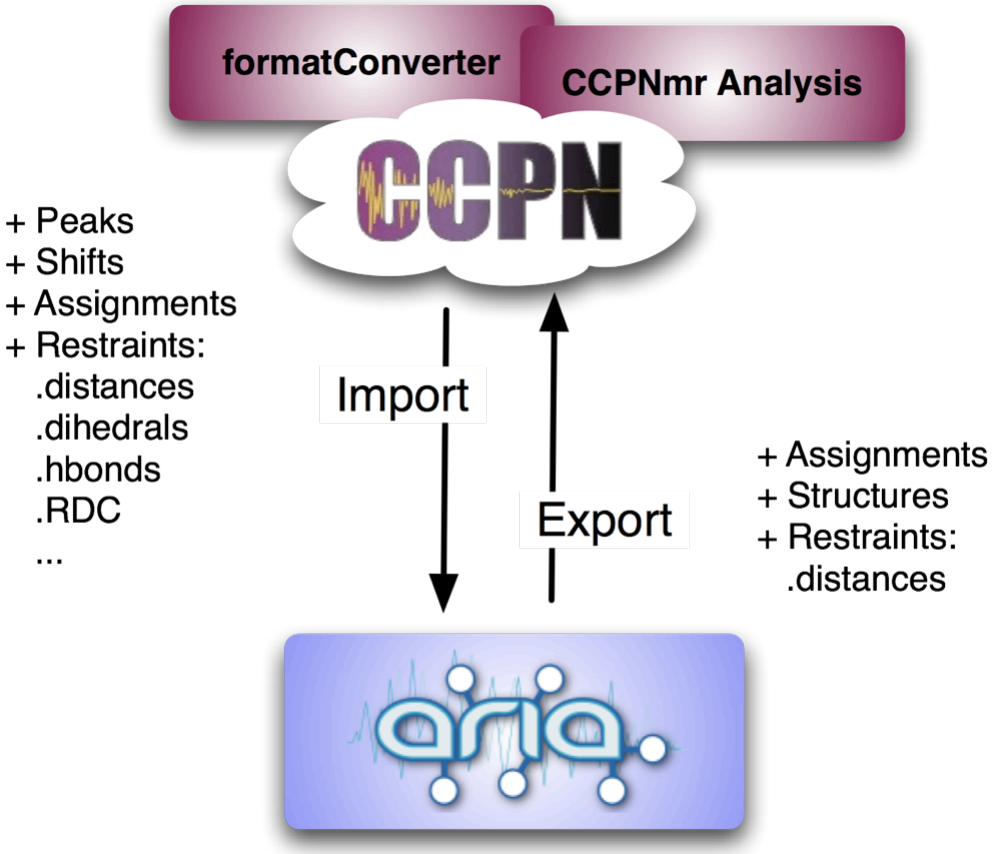
distance restraints
RMS deviation

GB1 tetramer (Gronenborn)



- Note: 6-stranded beta sheet with strands from each monomer
- Difficult; present limit of the method

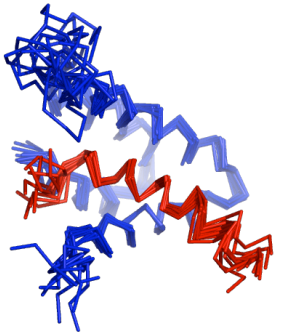
ARIA2.2 and complexes



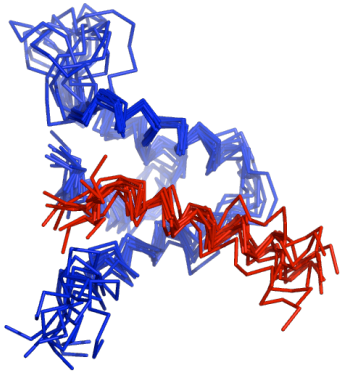
http://www.ccpn.ac.uk
Vranken *et al.* 2005

PAH2-SID24 complex
(van Ingen *et al.* 2004, 1PD7)

ARIA 2.2 structures
with original assignments



ARIA 2.2 structures
without original assignments

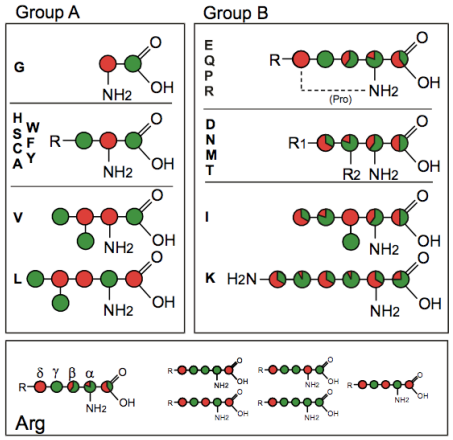


Calculation support for ssNMR

- Specific difficulties:
  - differnt type of distance data, distance dependency of signal
  - broader peaks, more ambiguities
  - additional inter-molecular contacts (in filament, microcrystals)
  - labelling schemes may require adapted assignment rules
- two types of experiments supported in ARIA
  - 2/3D 1H-driven 13C spin diffusion (PDSD)
  - 2D 1H spin diffusion (CHHC/NHHC)

ssNMR with ARIA: PDSD approach

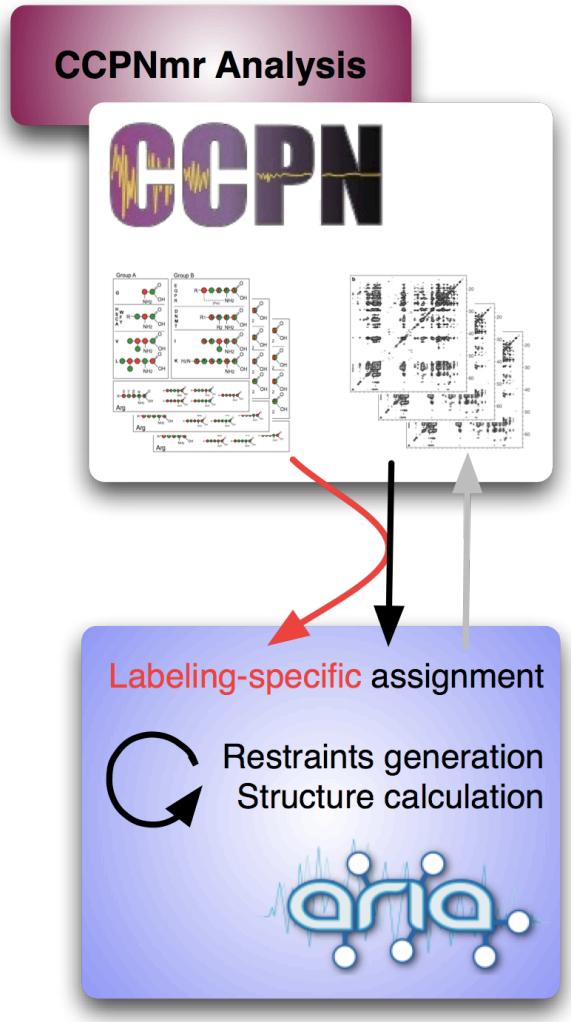
- CC correlations
- selectively labeled glycerol: 1,3- 13 C or/and 2-13C
- previous setup: SOLARIA (aria1.2 + ssNMR specific, Fossi et al. 2005)
  - assignments weighted according to labeling rates
  - loose bonds (~6.5 A)
  - SH3 structure
    - Coll. with B. van Rossum, H. Oschkinat (FMP Berlin)



SH3 strcuture caclcuated with SOLARIA
Precision 1.3A

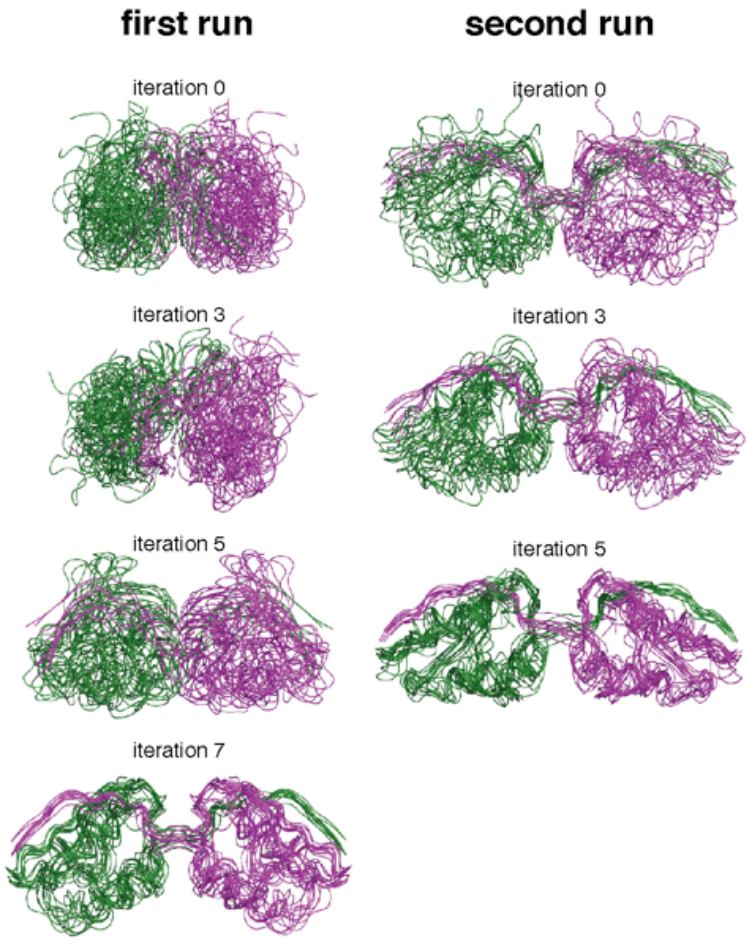
ssNMR with ARIA: PDSD approach

- ARIA2.2 already supports CCPN data exchange
  - add labeling schemes management and labeling specific assignment
  - reintroduce SOLARIA's features into ARIA 2
  - Collaboration with CCPN team (E. Laue, Cambridge) and H. Oschkinat group



ssNMR : uniformly labeled approach

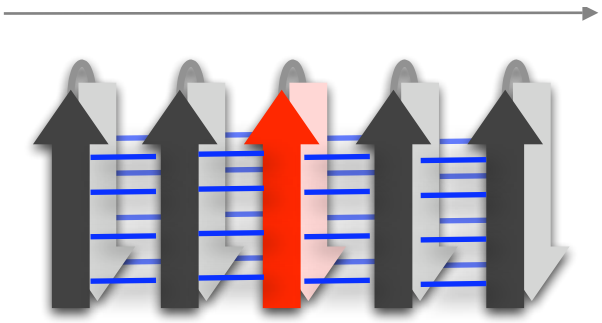
- Assignment and structure calculation with ARIA 2.2
- 1002 peaks, 2x16.3 assignment possibilities per peak
- 643 (65%) unambiguously assigned, 131 long-range
- N/C assignments but H-H restraints (1.8 - 5.0 A)
- coll. with A. Bockmann (IBPC Lyon)



Loquet A, Bardiaux B, Gardiennet C, Baldus ,M, Nilges M, Malliavin T and Bockmann A *JACS (2008)*

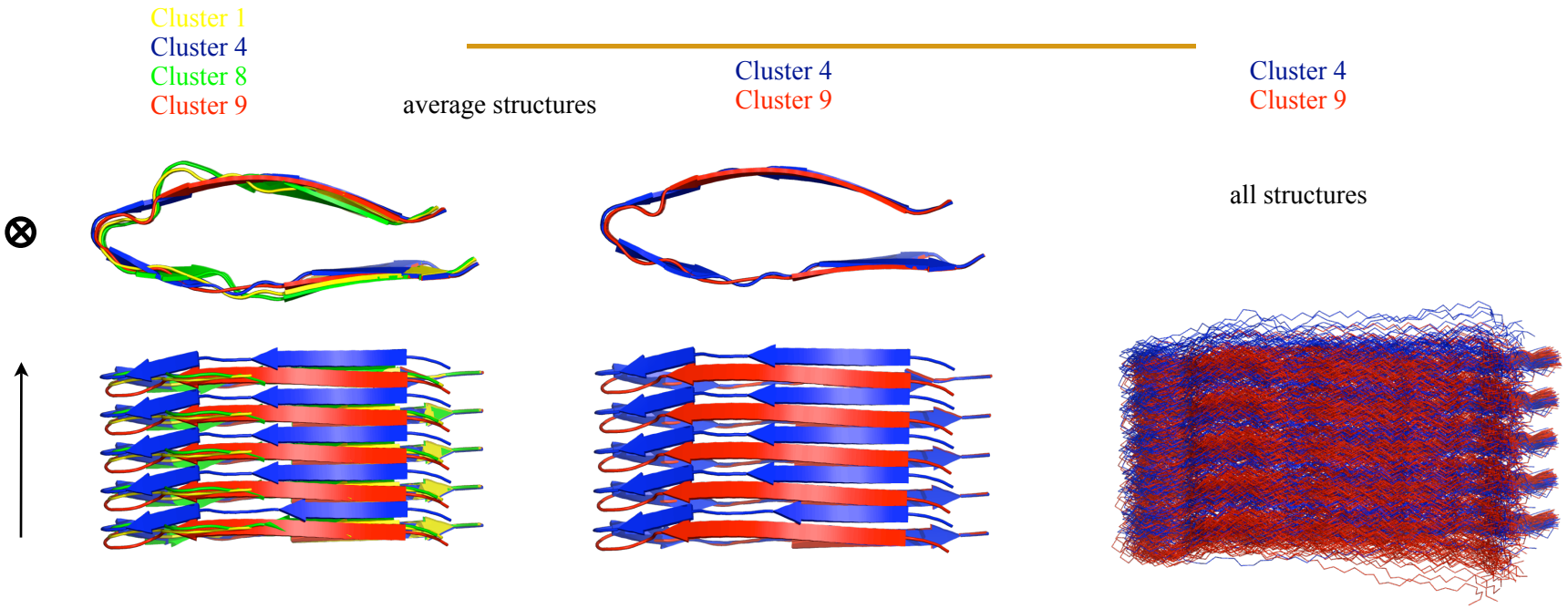
WW2 fibril calculations

- 4.73 translation along the fibril axis
- 4 images (+2, -2)
  - Phi/psi dihedral restraints
  - 25 ambiguous distances restraints
    - C-C, up to 5.5/7.5A
  - inter-monomeric hydrogen bonds restraints
- 1000 structures generated



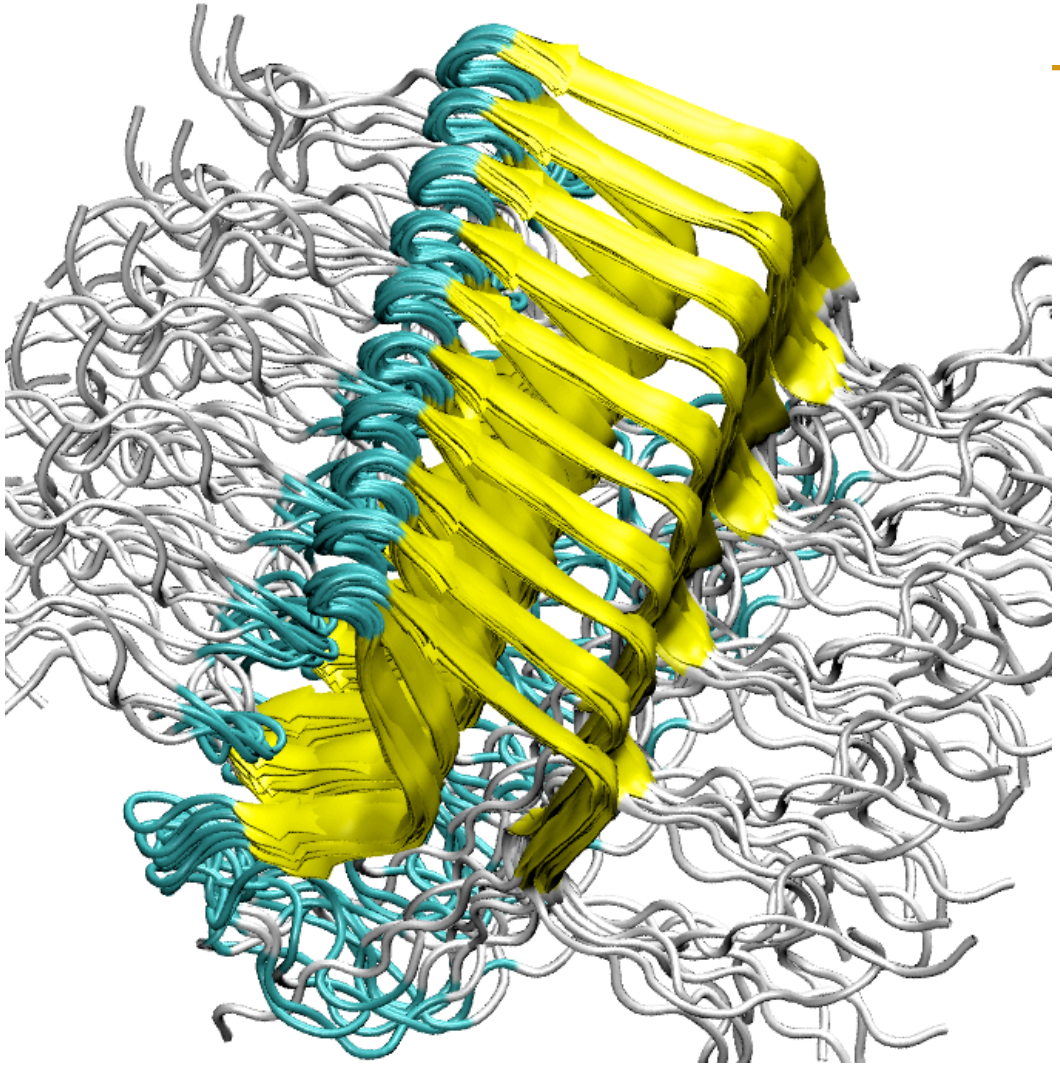
Central monomer
Images
Hydrogen bonds

Sheet alignment



- Average structures of clusters seems to show a clear distinction between the 2 possible alignments.
- No clear position when observing all the clustered structures

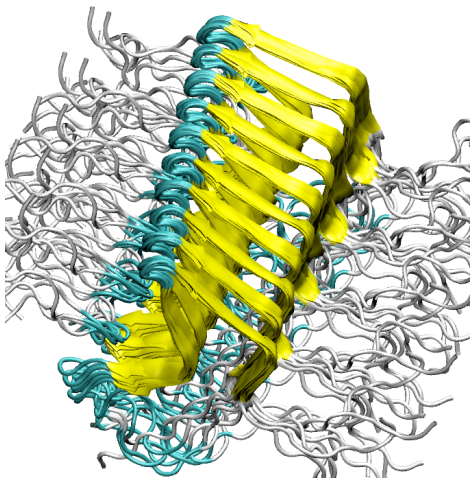
Fibril structures

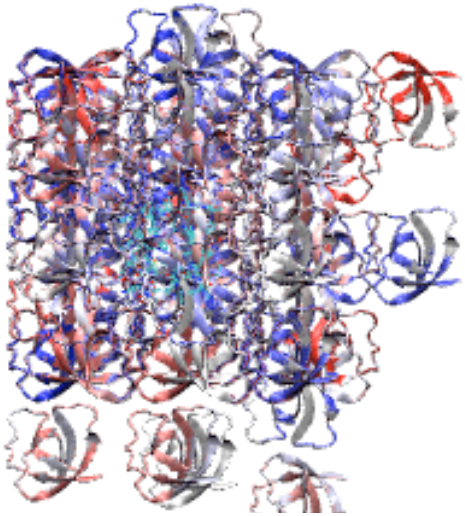


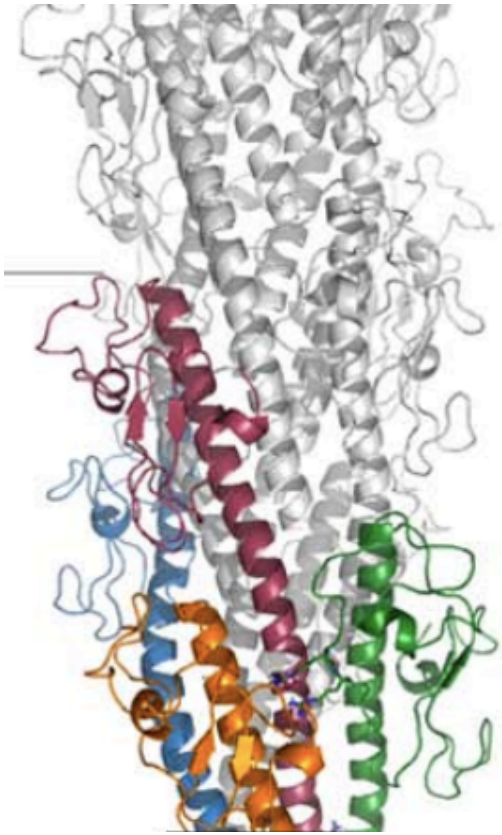
- With more complete data (from Riek et al., intra/ inter assignments removed)
- Note: twist is a calculation parameter

Other news in ARIA

- automatic superposition has always been part of it (Nilges, Gronenborn, Clore, FEBS lett 1987)
- truly general implementation of symmetry (all symmetries, including X-tal symmetries)
- solid state data sets
- soon: Bayesian determination of all weights
  - (coupling constants, RDCs)

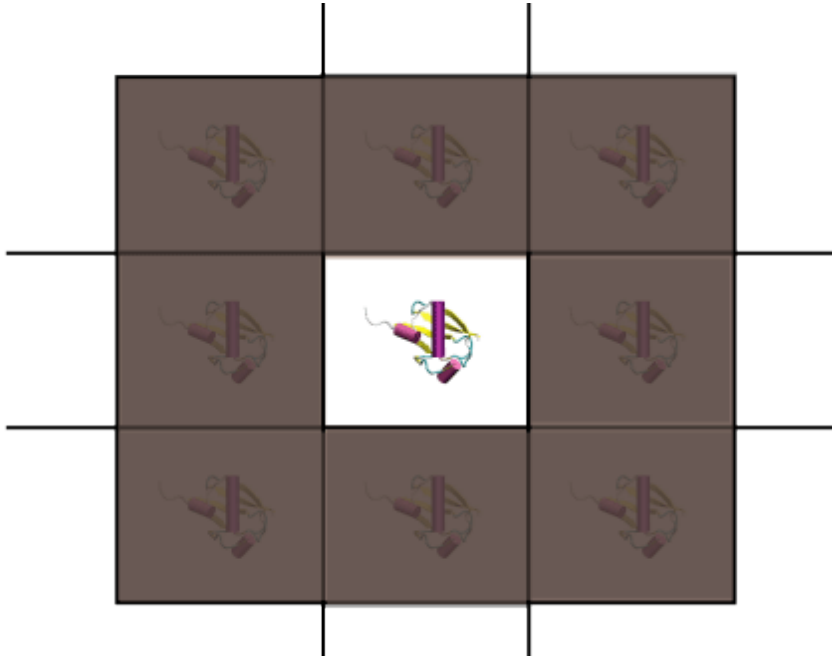




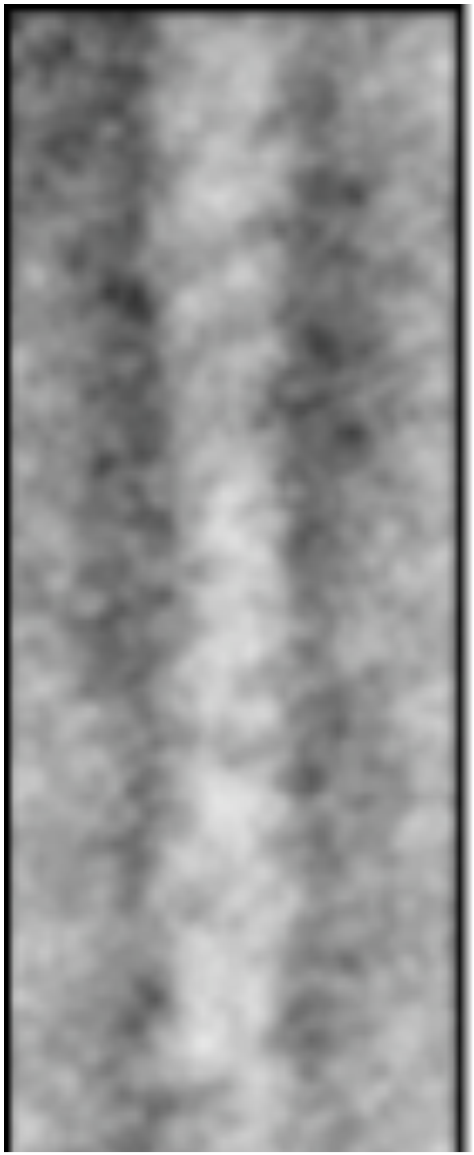


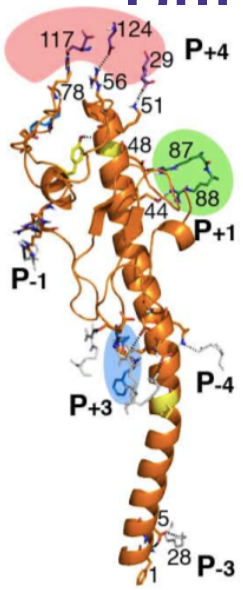
X-tal symmetries for ssNMR data

- flexible grid search with known structure
- folding (e.g., amyloid structures)



Low resolution EM + distances: structure of pilus

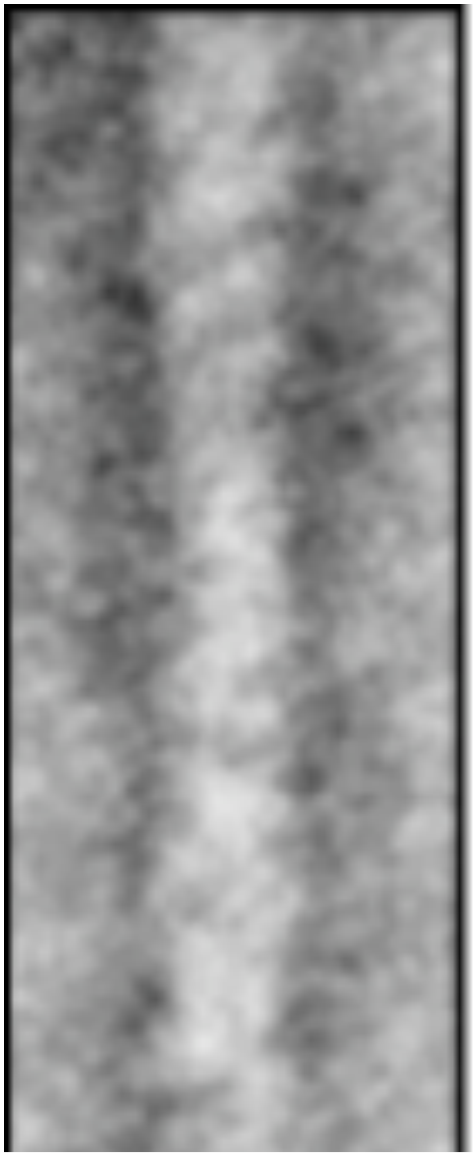


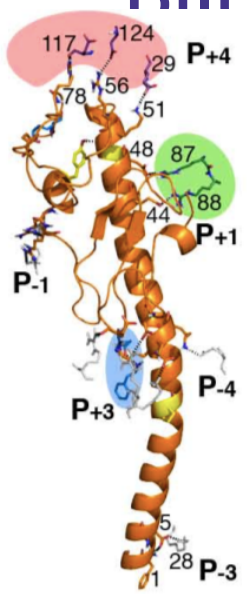


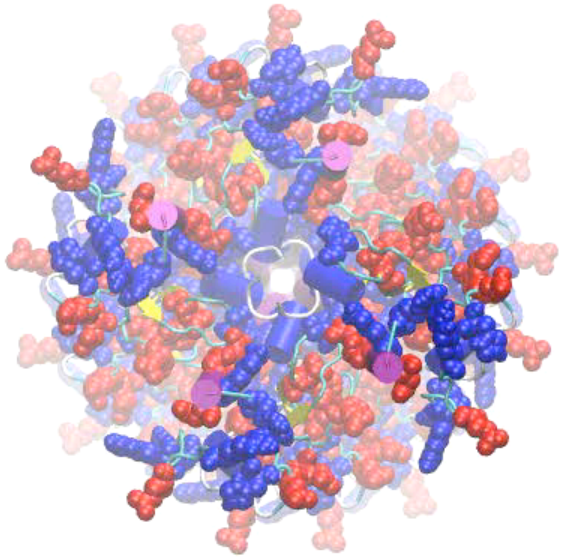
- X-ray structure, homology modelling
- symmetry (from EM)
- 2 distances (mutation experiments)

Campos et al., PNAS 2010

Low resolution EM + distances: structure of pilus



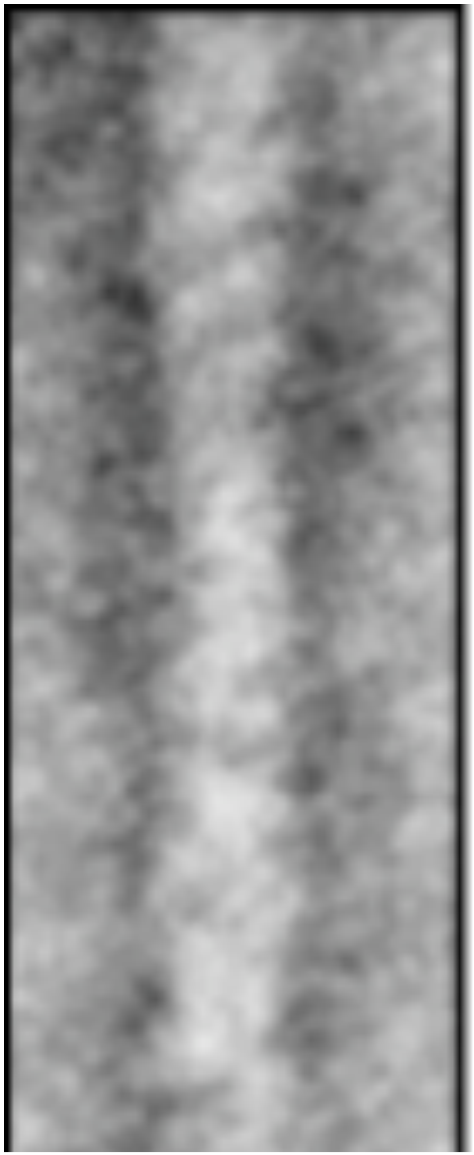


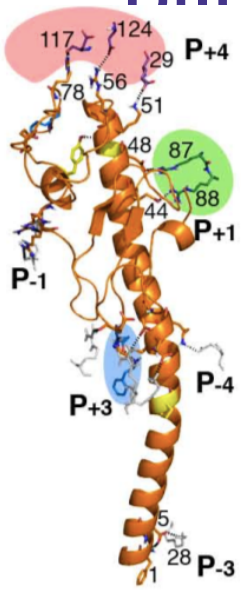


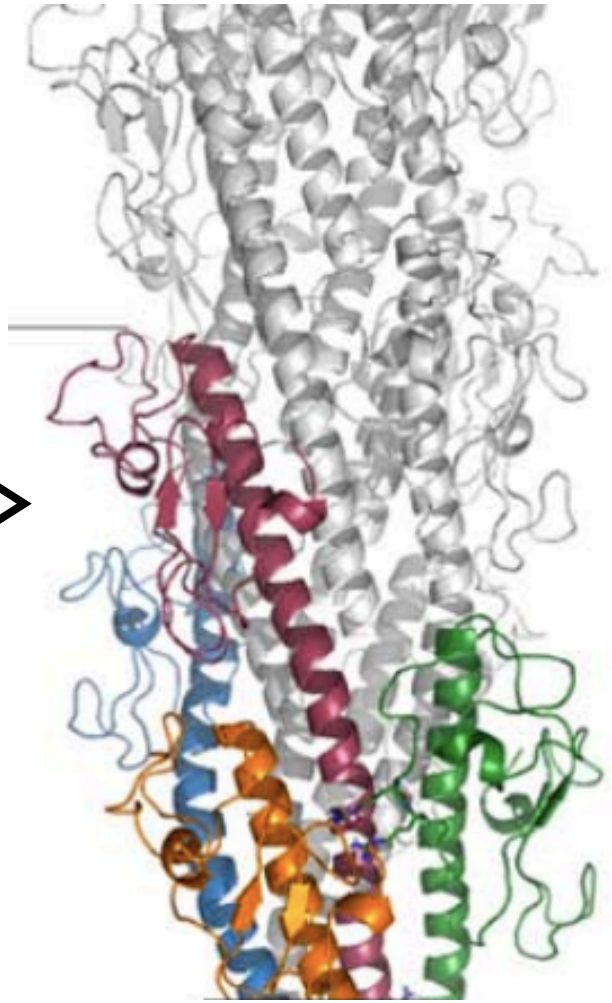
- X-ray structure, homology modelling
- symmetry (from EM)
- 2 distances (mutation experiments)

Campos et al., PNAS 2010

Low resolution EM + distances: structure of pilus







- X-ray structure, homology modelling
- symmetry (from EM)
- 2 distances (mutation experiments)

Campos et al., PNAS 2010

Overview

(1) Some basic concepts and history

(2) Recent developments

(3) Practical issues in ARIA2.2

WWW site: http://aria.pasteur.fr

site map accessibility contact

Search

home documentation download related links news

log in join

you are here: home

navigation

Home
Documentation
Download
Related links
News
Discussion group

log in

Login Name

Password

Log in

Forgot your password?

New user?

Welcome to ARIA

by wolfgang — last modified 2006-12-19 14:52

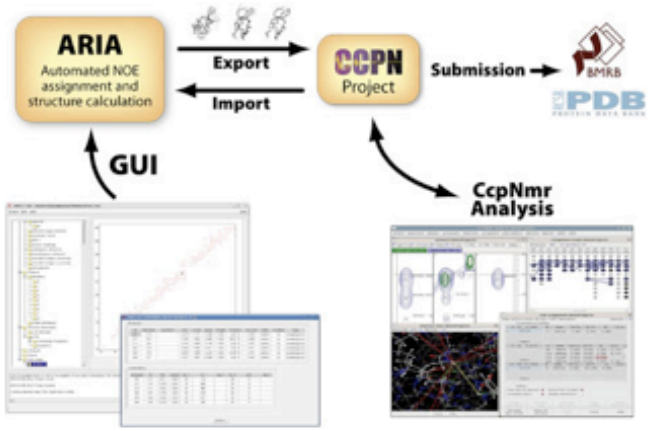
Our computer program ARIA (Ambiguous Restraints for Iterative Assignment) is a software for automated NOE assignment and NMR structure calculation. It speeds up and automatizes the assignment process through the use of an iterative structure calculation scheme. Additionally, a refinement in explicit water improves the quality of the calculated structures, validation tests help spectroscopists to judge the quality of the final structures, and the support of the CCPN data model simplifies the exchange of information with other NMR software packages.

May 2007

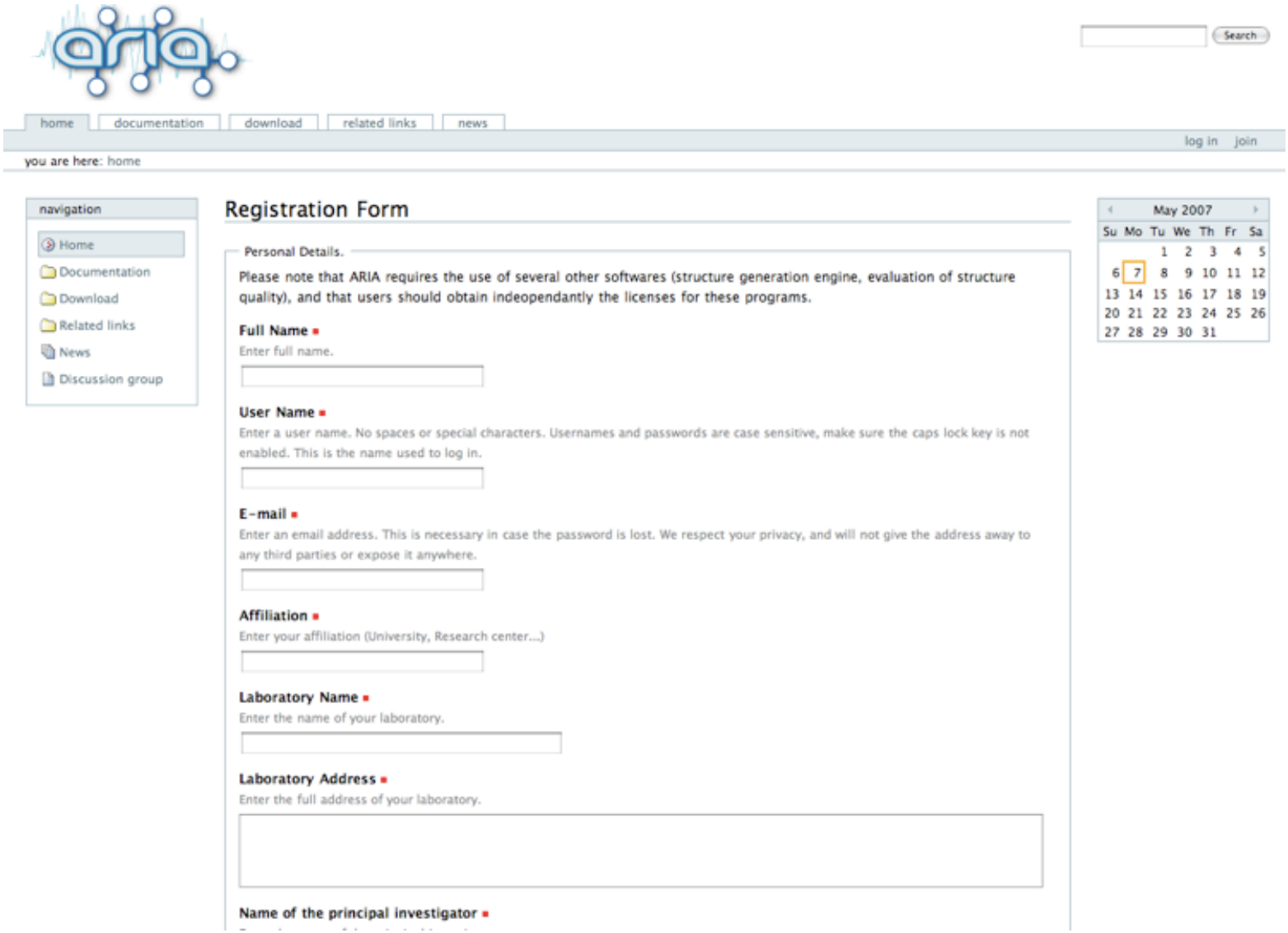
Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

How to obtain ARIA

ARIA comes with a free academic license. In order to download ARIA, you first need to register as a member of the ARIA web site. To do so, please click on the link "join", located at the top right which redirects you to the join page. After filling in the registration form, an e-mail is automatically sent to you. To activate your account, please follow the instructions given in the e-mail. Once you have logged in to the ARIA site, you can access the section "Download" located in the main navigation menu on the left.



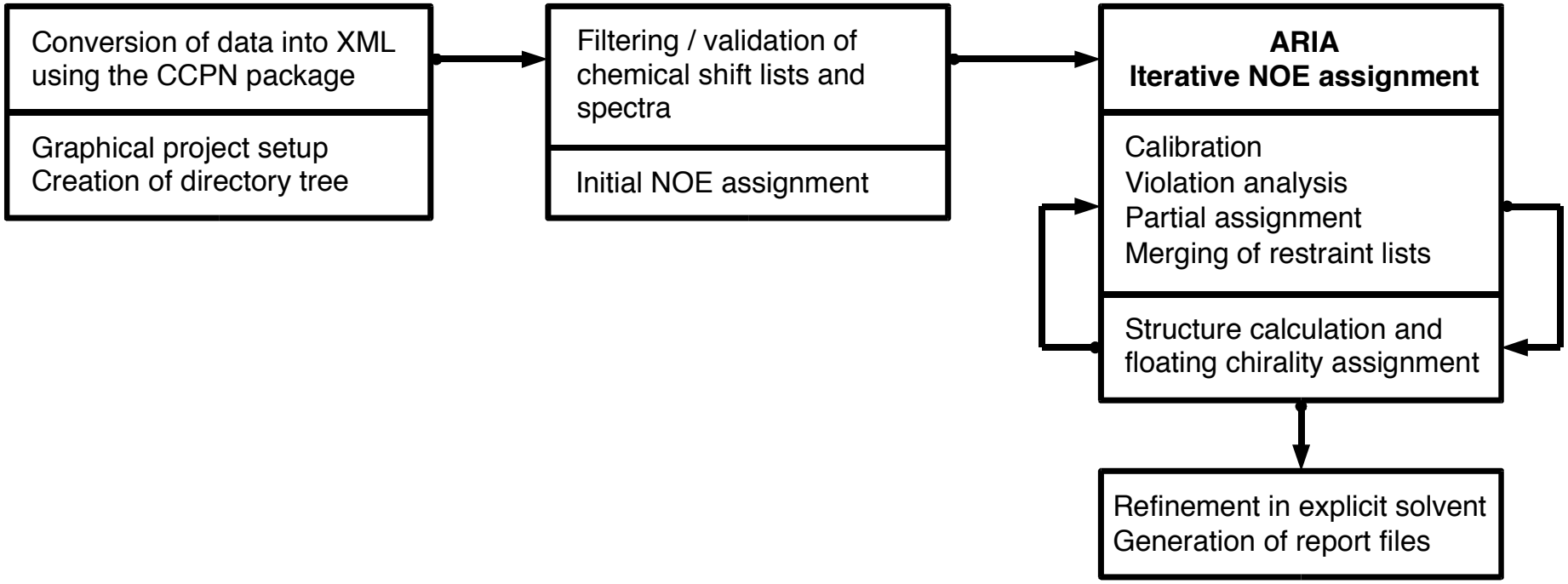
Registration form to download



ARIA2.2 program structure

- All data analysis in Python
- GUI in Tcl/Tk
- Structure calculation in CNS
- necessary CNS extensions (present or to migrate into Xplor-NIH)
  - floating chirality assignment
  - new developments (symmetry, log-harmonic potential)

ARIA2 program flow



ARIA2 input data

- Definition of the molecular system: AA sequence
- List of chemical shift assignments
- Peak list derived from NOESY spectrum
- Other experimental data that can be used by CNS:
  - scalar couplings,
  - residual dipolar couplings, diffusion anisotropy
  - PRE distance restraints,
  - SAXS,
  - dihedral angles,
  - hydrogen bond distance restraints

Data exchange with other software: CCPN

- Data model to store all results and analyses in general format
- CCPN project can be launched to start ARIA calculation
- Direct interaction with other programs complying with CCPN (e.g., CCPN Analysis)
- Data can be converted from a variety of formats (using CCPN Format Converter)

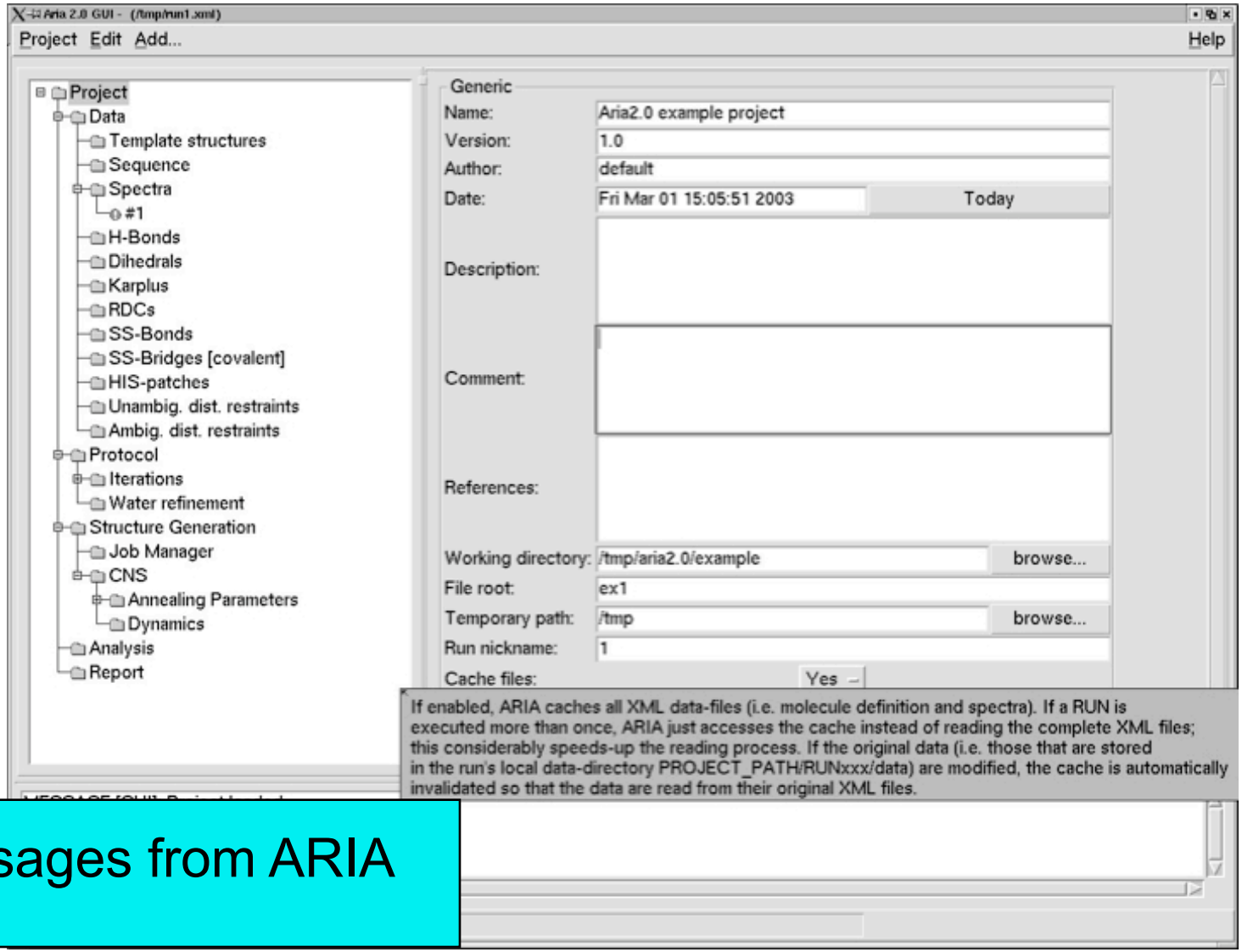
The project file

- Contains all program and protocol parameters
  - locations of the input data
  - program parameters
  - protocol parameters
  - generated automatically in data conversion
  - can also be created in GUI by choosing "New"
  - or by
  - `aria2 --project_template project.xml`
- Minimal description of molecule, NOESY peak list and chemical shift list
- Add more peak lists or other data
  - by "Add" in GUI

more project file mandatory items

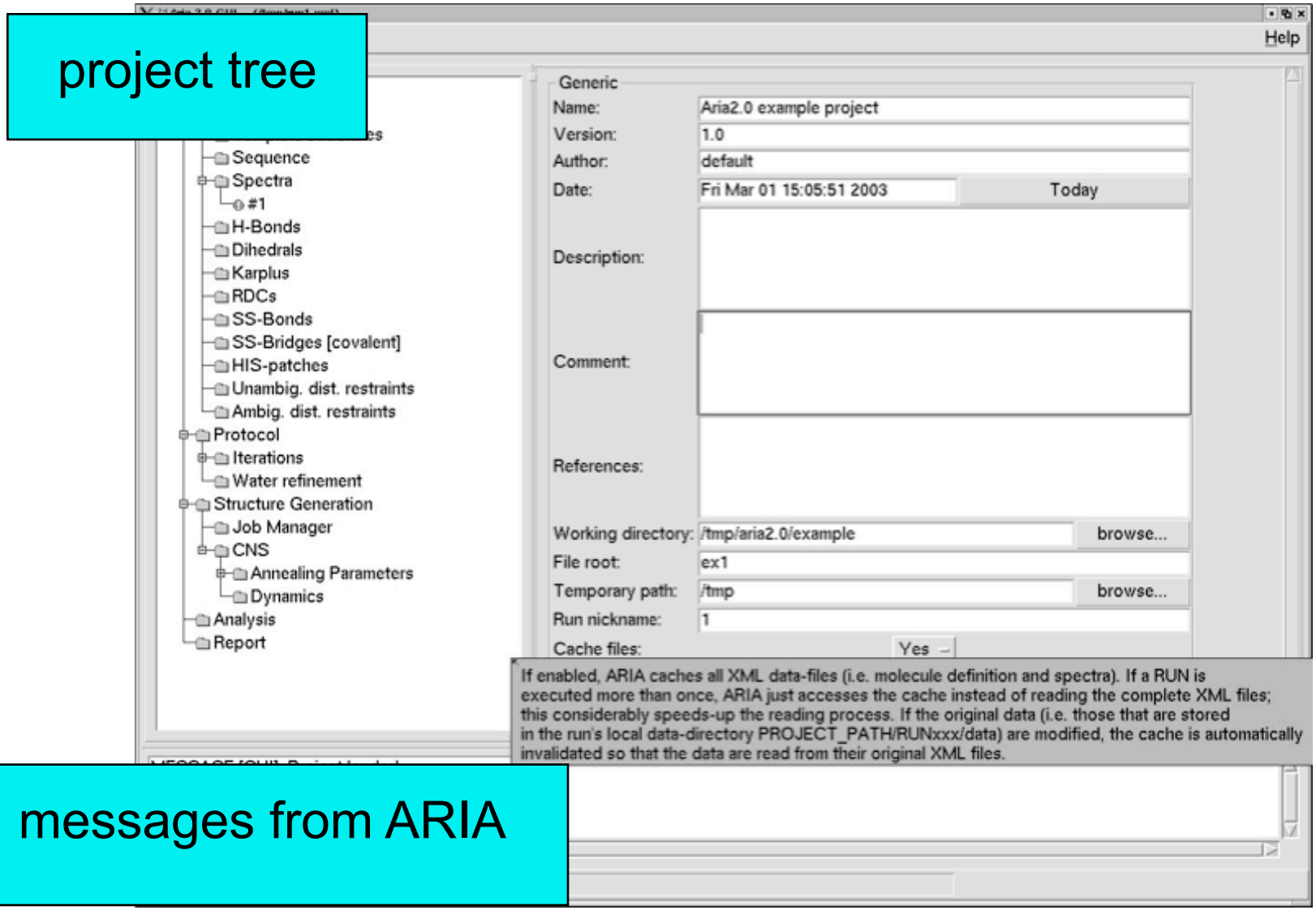
- Working directory
- File root (for naming output files)
- Temp directory for CNS
- CNS executable path

The GUI

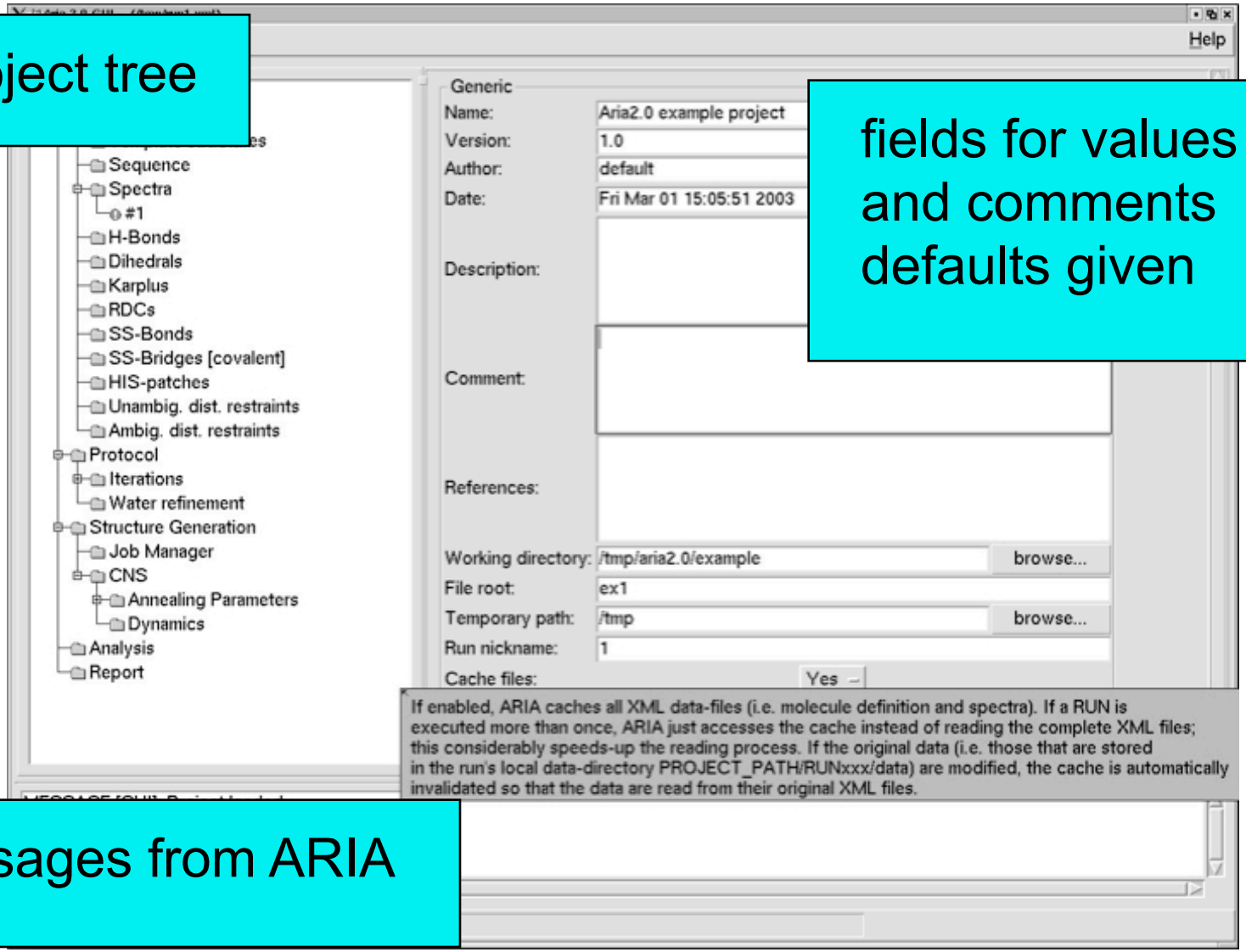


messages from ARIA

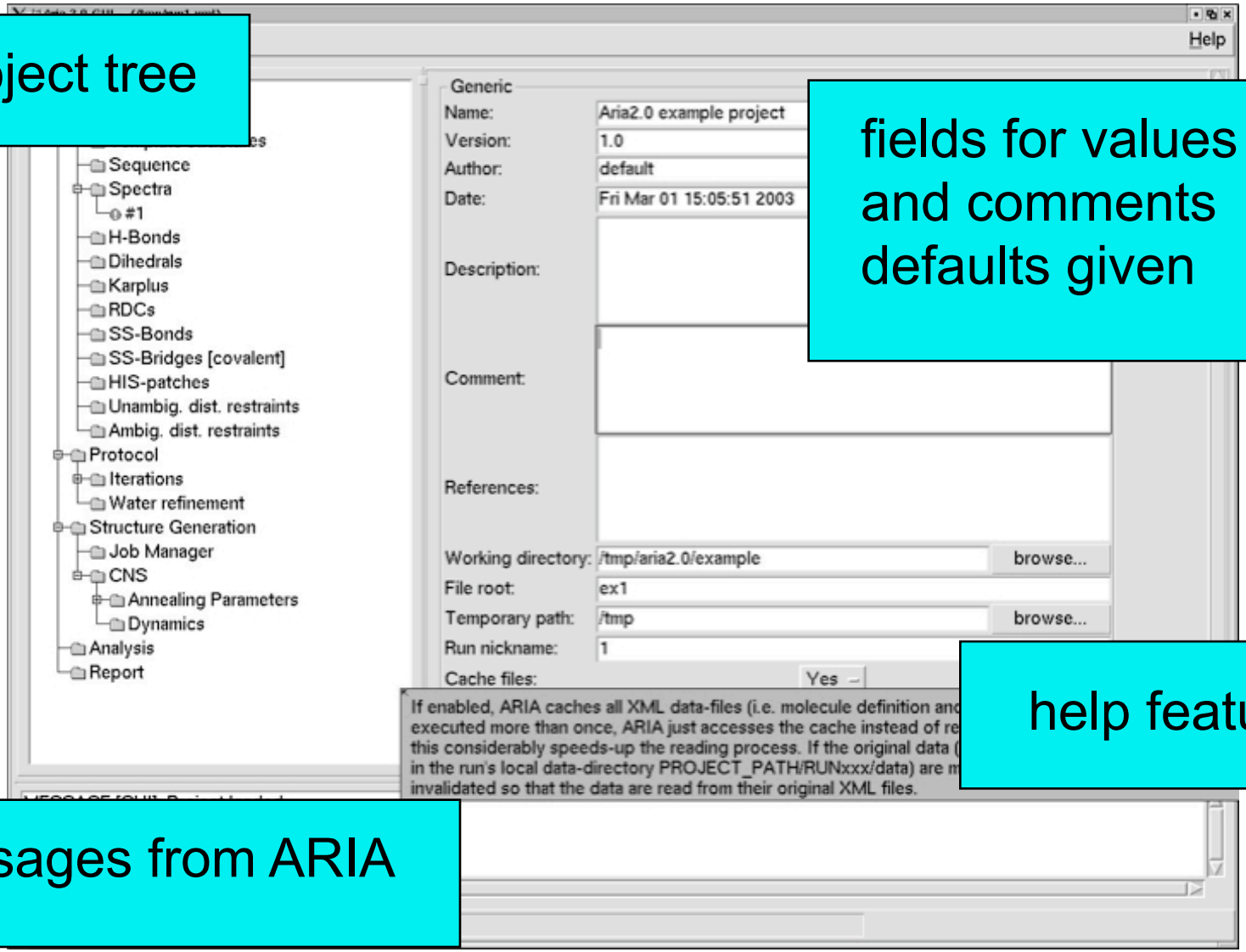
The GUI


project tree
Help
Sequence
Spectra
#1
H-Bonds
Dihedrals
Karplus
RDCs
SS-Bonds
SS-Bridges [covalent]
HIS-patches
Unambig. dist. restraints
Ambig. dist. restraints
Protocol
Iterations
Water refinement
Structure Generation
Job Manager
CNS
Annealing Parameters
Dynamics
Analysis
Report
Generic
Name:
Aria2.0 example project
Version:
1.0
Author:
default
Date:
Fri Mar 01 15:05:51 2003
Today
Description:
Comment:
References:
Working directory:
/tmp/aria2.0/example
browse...
File root:
ex1
Temporary path:
/tmp
browse...
Run nickname:
1
Cache files:
Yes
If enabled, ARIA caches all XML data-files (i.e. molecule definition and spectra). If a RUN is executed more than once, ARIA just accesses the cache instead of reading the complete XML files; this considerably speeds-up the reading process. If the original data (i.e. those that are stored in the run's local data-directory PROJECT_PATH/RUNxxx/data) are modified, the cache is automatically invalidated so that the data are read from their original XML files.
messages from ARIA

The GUI

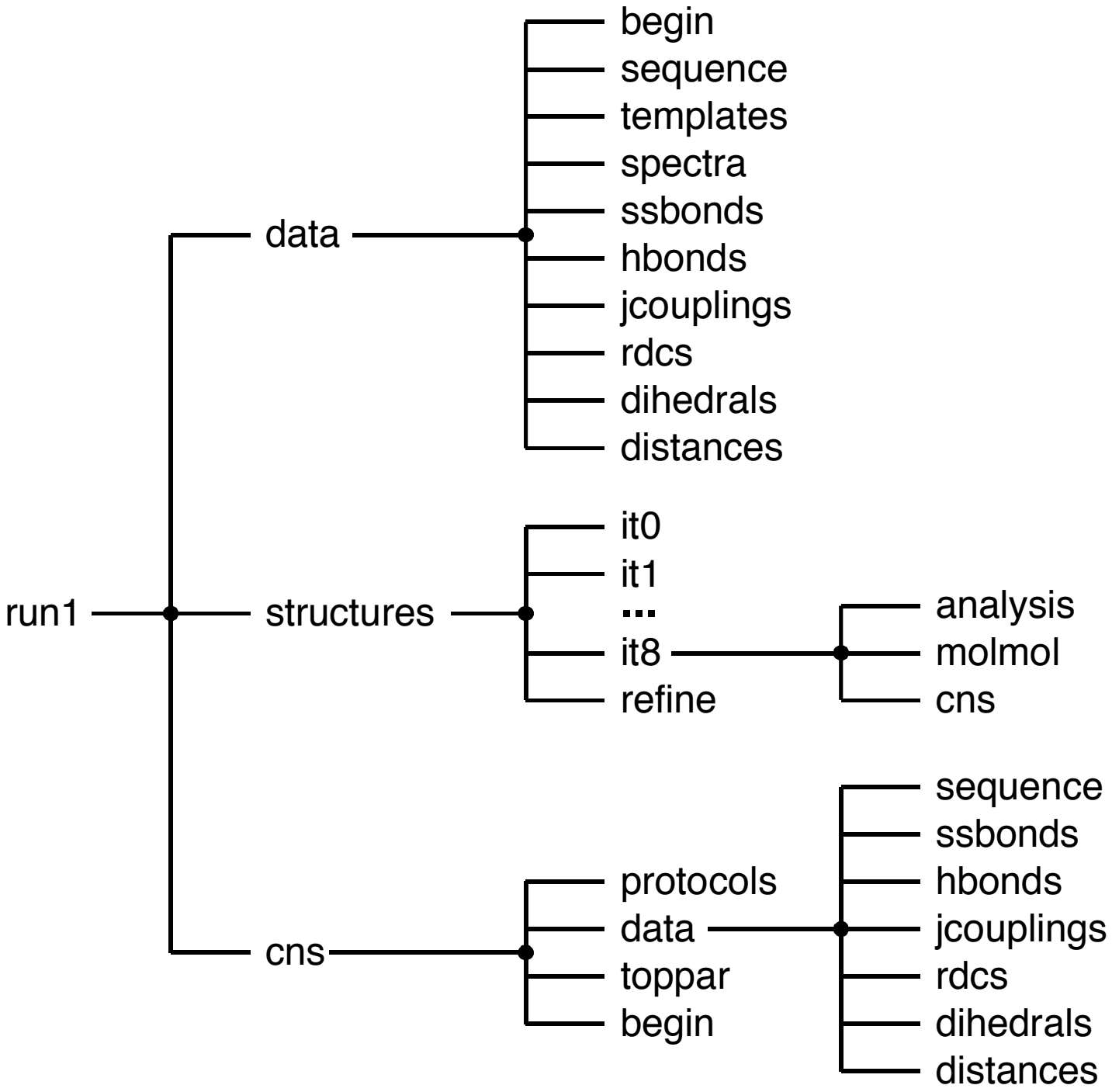

project tree
fields for values and comments defaults given
messages from ARIA
Help
Generic
Name:
Aria2.0 example project
Version:
1.0
Author:
default
Date:
Fri Mar 01 15:05:51 2003
Description:
Comment:
References:
Working directory:
/tmp/aria2.0/example
browse...
File root:
ex1
Temporary path:
/tmp
browse...
Run nickname:
1
Cache files:
Yes
If enabled, ARIA caches all XML data-files (i.e. molecule definition and spectra). If a RUN is executed more than once, ARIA just accesses the cache instead of reading the complete XML files; this considerably speeds-up the reading process. If the original data (i.e. those that are stored in the run's local data-directory PROJECT_PATH/RUNxxx/data) are modified, the cache is automatically invalidated so that the data are read from their original XML files.
Sequence
Spectra
#1
H-Bonds
Dihedrals
Karplus
RDCs
SS-Bonds
SS-Bridges [covalent]
HIS-patches
Unambig. dist. restraints
Ambig. dist. restraints
Protocol
Iterations
Water refinement
Structure Generation
Job Manager
CNS
Annealing Parameters
Dynamics
Analysis
Report

The GUI


project tree
fields for values and comments defaults given
help feature
messages from ARIA
Help
Generic
Name:
Aria2.0 example project
Version:
1.0
Author:
default
Date:
Fri Mar 01 15:05:51 2003
Description:
Comment:
References:
Working directory:
/tmp/aria2.0/example
browse...
File root:
ex1
Temporary path:
/tmp
browse...
Run nickname:
1
Cache files:
Yes
Sequence
Spectra
#1
H-Bonds
Dihedrals
Karplus
RDCs
SS-Bonds
SS-Bridges [covalent]
HIS-patches
Unambig. dist. restraints
Ambig. dist. restraints
Protocol
Iterations
Water refinement
Structure Generation
Job Manager
CNS
Annealing Parameters
Dynamics
Analysis
Report
executed more than once, ARIA just accesses the cache instead of re
this considerably speeds-up the reading process. If the original data (
in the run's local data-directory PROJECT_PATH/RUNxxx/data) are m
invalidated so that the data are read from their original XML files.

Setting up a new project

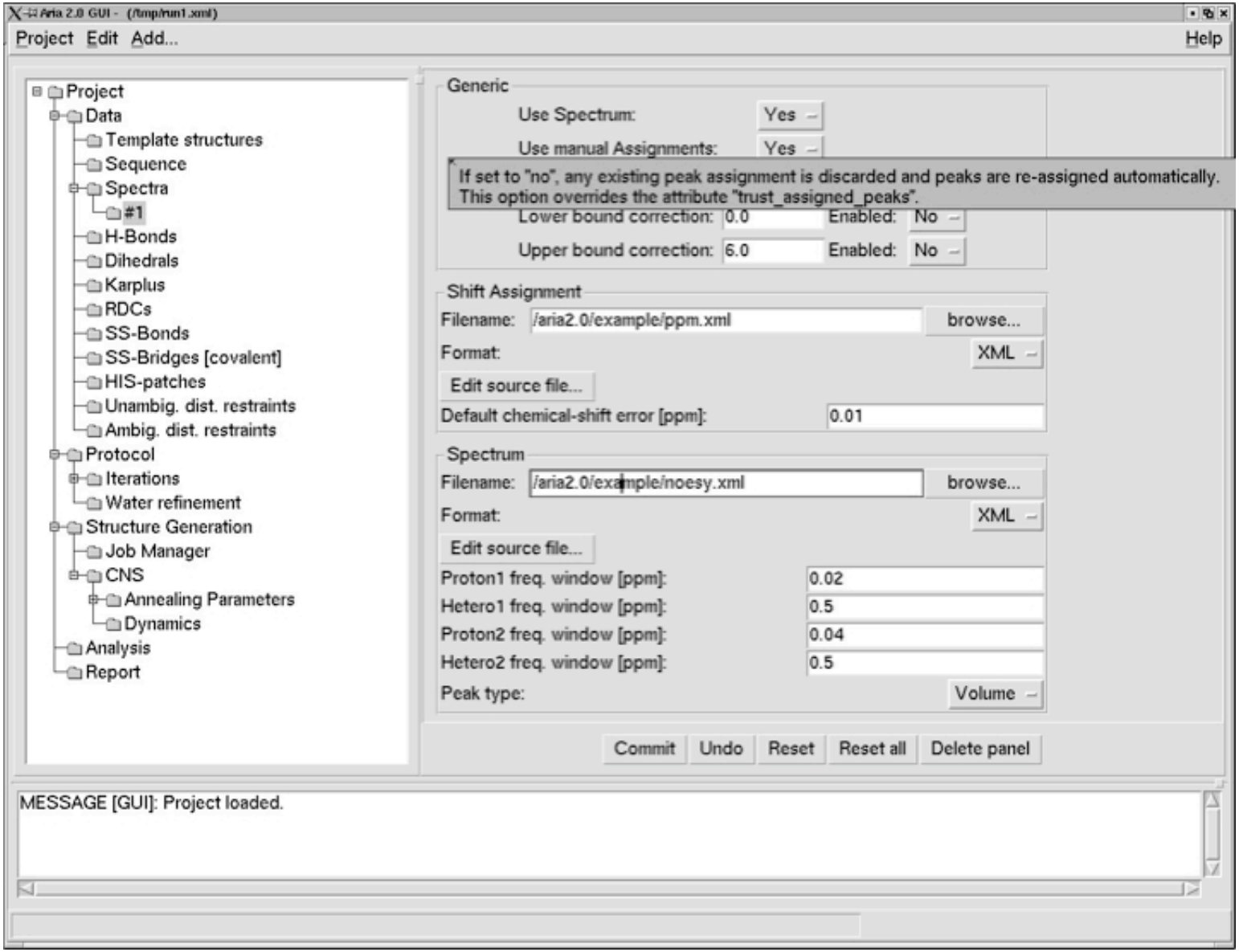
- Once the project file is ready, type
  - aria2 --setup project.xml
- this will generate a directory tree
  - primary data
  - structures with
    - iterations
    - all analysis results
  - cns for
    - protocols
    - CNS formatted data


run1
data
begin
sequence
templates
spectra
ssbonds
hbonds
jcouplings
rdcs
dihedrals
distances
structures
it0
it1
...
it8
refine
analysis
molmol
cns
cns
protocols
data
toppar
begin
sequence
ssbonds
hbonds
jcouplings
rdcs
dihedrals
distances

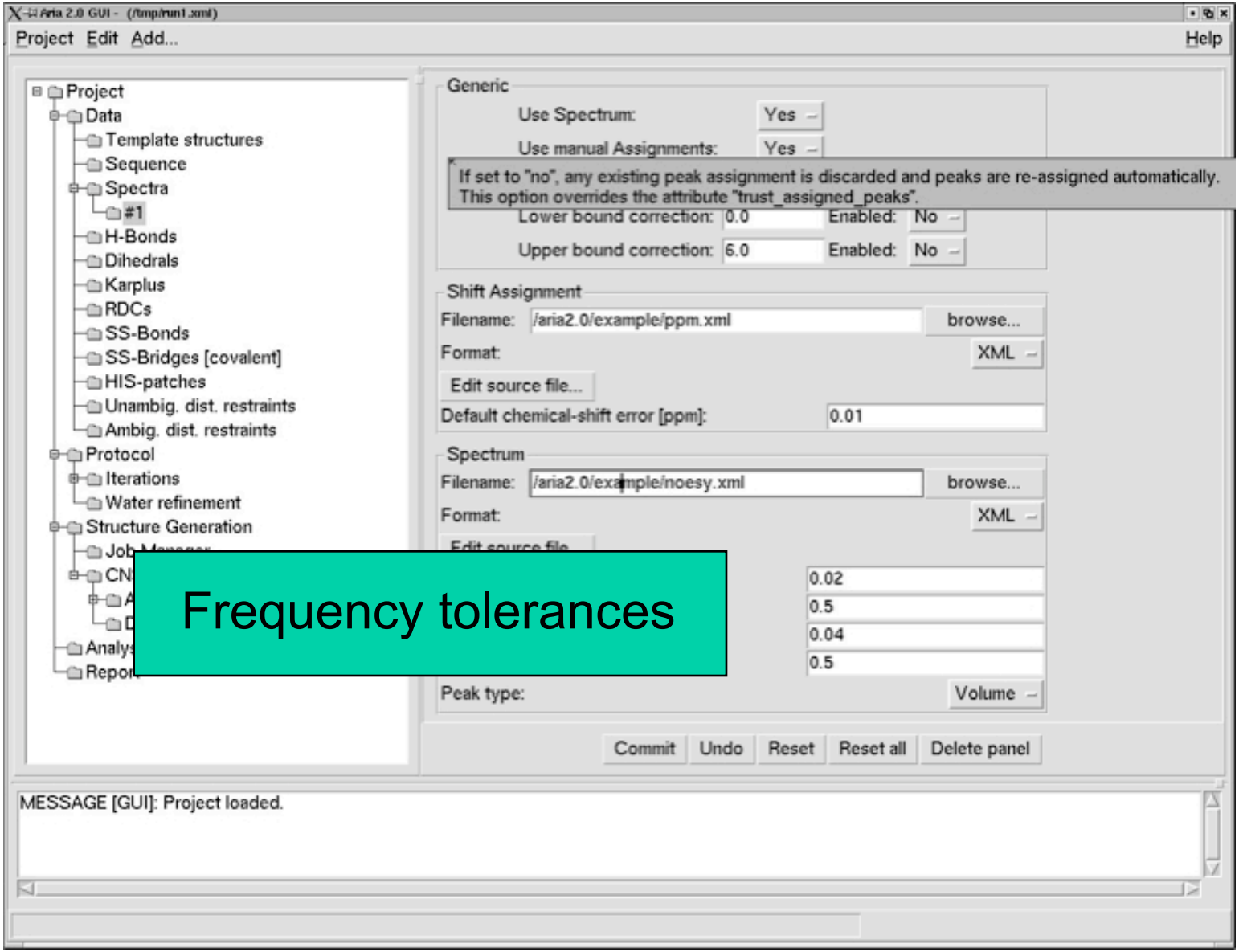
Preparation stage

- Filtering of data for errors and inconsistencies
  - chemical shifts: unique, degenerate (motion), prochiral
  - peak list: removal of diagonal peaks, use of absolute values for intensities
- Creation of molecular topology file (MTF file)
  - CNS topology file derived from sequence using CNS script
  - pseudo-extended template file
  - standard biopolymers: automatic generation
- "Seed assignment" (list of all possible assignments purely based on chemical shift and frequencey tolerances

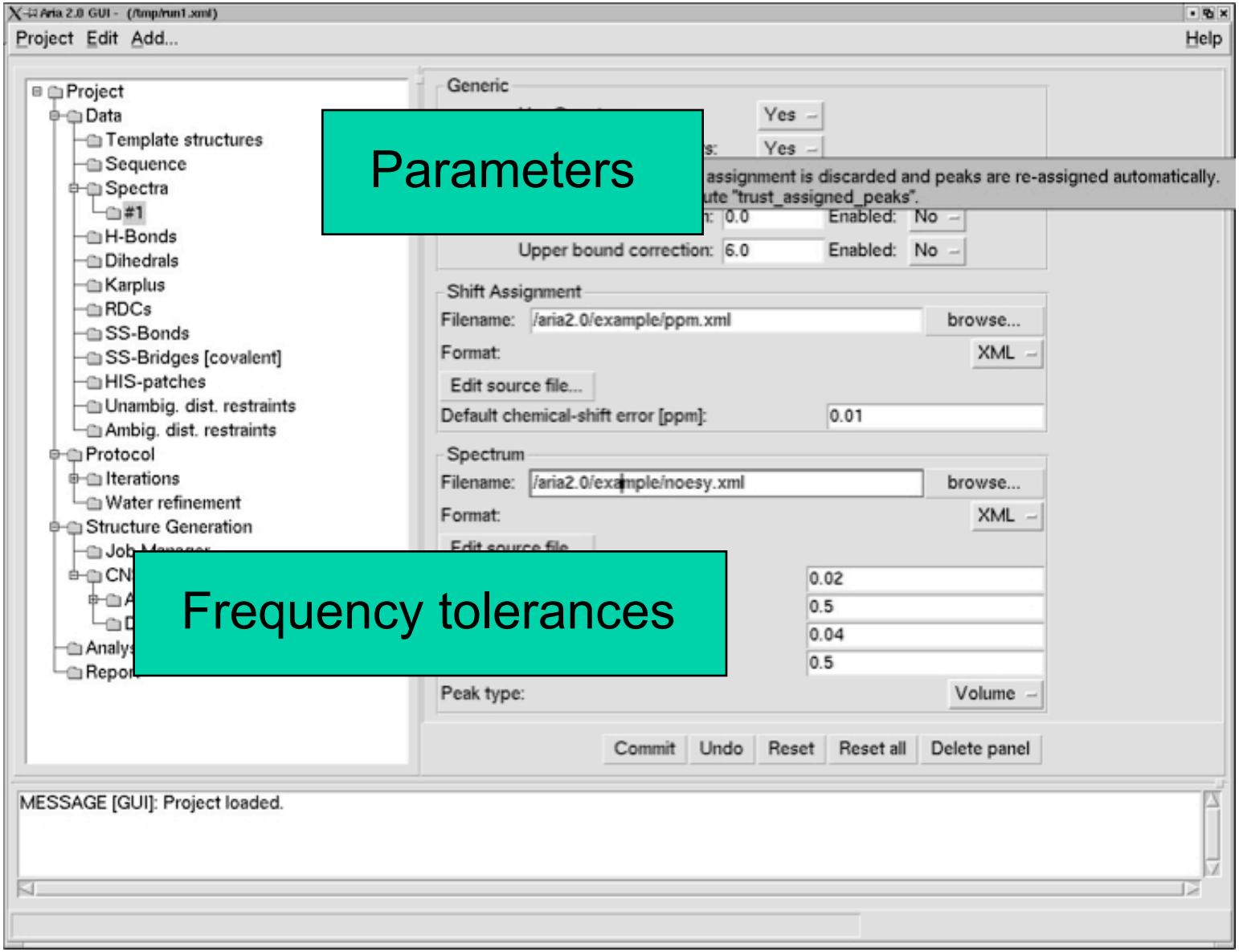
GUI for spectrum



GUI for spectrum



GUI for spectrum



Data conversion

- first action: convert all data to XML
- needs:
  - file names and locations
  - additional information: XML conversion file
  - generate a template file to edit by:
  - ```
    aria2 --convert -t conversion.xml
    ```
  - then convert the data by
  - ```
    aria2 --convert -t conversion.xml
    ```
  - this also generates a project file

Formats for data representation

- Data format based on XML:
  - portable
  - human readable
  - document validation
  - strict checking

```
167   2.051 0.001 HB2    10
168   1.970 0.001 HB3    10
169 999.000 0.000 QB     10
```

```
<shift_assignment method="FLOATING">
  <spin_system averaging_method="NONE">
    <atom segid="    " residue="10" name="HB2"/>
    <chemical_shift value="1.97" error="0.001"/>
    <chemical_shift value="2.051" error="0.001"/>
  </spin_system>
  <spin_system averaging_method="NONE">
    <atom segid="    " residue="10" name="HB3"/>
    <chemical_shift value="1.97" error="0.001"/>
    <chemical_shift value="2.051" error="0.001"/>
  </spin_system>
</shift_assignment>
```

Preparation of structure calculation

- Each spectrum is treated separately (leads sometimes to problems if there are very few peaks)
- Merge all *accepted* restraints into one distance restraint list by removing duplicates
- Always keep restraint with smallest bound difference U-L

Running ARIA

- To start ARIA:
  - aria2 project.xml
- recommended to use parallel setup
- project file must provide command to execute CNS
  - "Attribute" in GUI
  - e.g., `ssh hostname csh`

Interpreting the results

- Total energies should be below 1000 kCal/mol - depending on details of the calculation (e.g. dihedrals on or off)
- Large average energy, large energy variation
  - increase number of dynamic steps
  - increase bound width
- Low average energy, large energy variation
  - increase number of dynamic steps

Analysis and output files

- NOE files
  - noe_restraints.unambig, noe_restraints.ambig
  - noe_restraints.violations
  - noe_restraints.assignments
  - noe_restraints.merged
  - noe_restraints.xml
  - noe_restraints.pickle
  - report
- MOLMOL distance files for graphics
- Structure analysis with WHATIF, PROCHECK, PROSA
- miscellaneous CNS analysis output, including average

Checking rejected NOEs

ref_spec[1]	ref_no[1]	lower[2]	upper[2]	d_avg[3]	u_viol	%_viol	a_type[4]	n_c[5]
...								
13C NOESY	1568	2.03	5.00	14.73	9.73	100.0	S	2
13C NOESY	1612	1.76	3.30	12.93	9.63	100.0	M	1
13C NOESY	2996	1.89	3.90	13.38	9.49	100.0	A	3
...								

- rejected NOEs should be checked, in particular if the size of the violations

Acknowledgements

- IP Paris: T. Malliavin, F. Mareuil, A. Bernard, B. Bardiaux (now Berlin)
- ARIA2.0: M. Habeck, W. Rieping, J. Linge
- ARIA1.0: J. Linge, S. O'Donoghue

- CCPN: U. Cambridge/ EBI Hinxton: W. Vranken, T. Stevens, R. Fogh, W. Rieping, E. Laue
- Solid state: IBPC Lyon: A. Loquet, A. Böckmann + group; FMP Berlin: B. van Rossum, M. Fossi, H. Oschkinat + group