Supplementary Information

Impact of ^{15}N R_2/R_1 Relaxation Restraints on Molecular Size, Shape and Bond Vector Orientation for NMR Protein Structure Determination with Sparse Distance Restraints

Yaroslav Ryabov,[1] Charles D. Schwieters[1,*] and G. Marius Clore,[2,*]

[1]Division of Computational Bioscience, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624, and [2]Laboratory of Chemical Physics, Building 5, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

Details of structure determination protocol:

Step 1: Generation of extended strand conformation.
Step 2: Torsion angle randomization.
Step 3: First phase of high temperature torsion angle dynamics, using only distance (hydrogen bond and/or NOE) and ϕ/ψ torsion angle. Duration 100 ps.
Step 4: Second phase of high temperature torsion angle dynamics, using distance, torsion angle and ^{15}N-R_2/R_1 restraints. Duration 100 ps.
Step 4: Simulated annealing.
Step 5: Final gradient minimization first in torsion angle space and then in Cartesian space.

Simulated annealing:
Starting temperature 3500 K
Final temperature: 25 K
Temperature steps: 12.5 K
Duration of simulations at every temperature: 0.5 ps or 3000 steps of molecular dynamics, whichever happens first.

100 structures are calculated and the 10 lowest energy structures are analyzed.

Table S1 Potential terms employed

Potential term (unit of force constant)	Restraints	Force constant			
		High temperature dynamics		Simulated annealing	Final minimization
		Phase I	Phase II		
`relaxRatioPot`[a] (kcal/mol)	ratios of R_2/R_1 relaxation rates	0	0.001	ramped from 0.01 to 1	1
`NOEpot` (kcal/mol/Å^2)	square-well distance restraints from NOEs and hydrogen bonds	2	20	ramped from 20 to 100	100
`CDIH` (kcal/mol/rad^2)	square-well torsion angle restraints	500	10,000	10,000	10,000
`VDW` (kcal/mol/Å^4)	quartic atom-atom repulsion	0.01 Cα atoms	0.01 Cα atoms	ramped from 0.01 to 4 for all atoms	4
`BOND` (kcal/mol/Å^2)	bond length	1000	1000	1000	1000
`ANGLE` (kcal/mol/rad^2)	bond angle	200	200	ramped from 200 to 500	500
`IMPR` (kcal/mol/rad^2)	improper dihedral	50	50	ramped from 50 to 500	500

[a]In initial high temperature dynamics the effective diffusion tensor temperature T_{diff}^{app} (which is not a physical temperature but a fitting parameter which includes the effect of actual temperature and viscosity) was fixed to the nominal experimental temperatures of 297, 300 and 316K for GB3, ubiquitin and EIN, respectively. During simulated annealing T_{diff}^{eff} was optimized within a window that was geometrically increased from ~0 to ±10 K.

Python script for structure determination protocol used in Xplor-NIH

```
# relaxation restraints for protein structure determination
#
# Y. Ryabov, C.D. Schwieters, and G.M. Clore (2011)
# set up initial parameters of the script
xplor.requireVersion("2.18")

outFilename = "SCRIPT_STRUCTURE.sa"
numberOfStructures=100

import protocol
protocol.initRandomSeed(3421) #explicitly set random seed

command = xplor.command

protocol.initParams("protein")

# generate PSF data from sequence
from psfGen import seqToPSF
seqToPSF('ub.seq')

# generate initial extended strand structure
protocol.genExtendedStructure()

from potList import PotList
potList = PotList()
potListIn = PotList()

from simulationTools import MultRamp, StaticRamp, InitialParams

rampedParams=[]
highTempParams=[]
highTempParams2=[]

# set up the energy term that use 15N NMR relaxation data
from diffPotTools          import readInRelaxData, mergeRelaxData
from relaxRatioPotTools import create_RelaxRatioPot

# In this particular case the relaxation data are provided in
# the form of separate files for the T1 and T2 data,
# each of which comprises two different measurements.
# N. Tjandra, S.E. Feller, R.W. Pastor, and A. Bax
# JACS 117 (1995) 12562
# Initially the scripts reads data files
T1_1 = readInRelaxData(['T1_1_600'], pattern=['skip','resid','T1'])
T1_2 = readInRelaxData(['T1_2_600'], pattern=['skip','resid','T1'])

# The experimental data do not have explicit
# values of experimental error.
# Original paper reports about %0.5 in measured relaxation data
# However our simulations suggest a more realistic value of 2%
# thus the following lines set up proper values of errors for T1 data
for item in T1_1:
   item.R1_err=0.02*item.R1
for item in T1_2:
   item.R1_err=0.02*item.R1

# this command merges together two sets of T1 data into a single
# list of mean T1 values with properly weighted errors
```

```
T1_600=mergeRelaxData(T1_1+T1_2)

# reading in T2 data
T2_1 = readInRelaxData(['T2_1_600'], pattern=['skip','resid','T2'])
T2_2 = readInRelaxData(['T2_2_600'], pattern=['skip','resid','T2'])

# setting errors for T2
for item in T2_1:
   item.R2_err=0.02*item.R2
for item in T2_2:
   item.R2_err=0.02*item.R2

# merging two sets of T2 values
T2_600=mergeRelaxData(T2_1+T2_2)

# combining together T1 and T2 data into R2/R1 ratios
# with properly weighted errors
data_600=mergeRelaxData(T1_600+T2_600)

# this example is specific to the particular set
# of relaxation data being used. Xplor-NIH, however, accepts
# a large variety of relaxation data input formats.
# For further details read specifications for the
# readInRelaxData routine in diffPotTools.py
#
# creating the relaxRatioPot potential instance
r_600=create_RelaxRatioPot('rrp_600', data_in = data_600,
    freq = 600, temperature = 273+27, addAtoms=True)

# setting up the initial state of relaxRatioPot
p_tess=r_600
potList.append(r_600)
r_600.setScale(0.001)
r_600.setRangeTmpFit(0)
r_600.setSigmaFactor(1.5)

# setting up parameters of ramping for relaxRatioPot
highTempParams.append( StaticRamp("r_600.setScale( 0.001 )") )
highTempParams2.append( StaticRamp("r_600.setScale( 0.001 )") )
rampedParams.append( MultRamp(0.01,1, "r_600.setScale( VALUE )") )
rampedParams.append( MultRamp(0.0,10.0, "r_600.setRangeTmpFit( VALUE )") )

# set up NOE potential
# in this case NOEPot is used to provide

# pseudo NOE distance restraints
# derived from hydrogen bond connectivity only
noe=PotList('noe')
potList.append(noe)
potListIn.append(noe)
from noePotTools import create_NOEPot
for (name,scale,file) in [ ('hb',1,"H_bonds")]:
   pot = create_NOEPot(name,file)
   # pot.setPotType("soft") - if you think there may be bad NOEs
   pot.setScale(scale)
   noe.append(pot)
```

```
highTempParams.append( StaticRamp("noe.setScale( 2 )") )
highTempParams2.append( StaticRamp("noe.setScale( 20 )") )
rampedParams.append( MultRamp(20,100, "noe.setScale( VALUE )") )

from xplorPot import XplorPot

#Rama torsion angle database
#
protocol.initRamaDatabase()
potList.append( XplorPot('RAMA') )
potListIn.append( potList['RAMA'] )
rampedParams.append( MultRamp(.002,1,"potList['RAMA'].setScale(VALUE)") )

# setup parameters for atom-atom repulsive term. (van der Waals-like term)
potList.append( XplorPot('VDW') )
potListIn.append( potList['VDW'] )
rampedParams.append( StaticRamp("protocol.initNBond()") )
rampedParams.append( MultRamp(0.9,0.8,
                    "command('param nbonds repel VALUE end end')") )
rampedParams.append( MultRamp(.004,4,
                    "command('param nbonds rcon VALUE end end')") )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("""protocol.initNBond(cutnb=100,
                                        rcon=0.004,
                                        tolerance=45,
                                        repel=0.9,
                                        onlyCA=1)""") )

highTempParams2.append( StaticRamp("""protocol.initNBond(cutnb=100,
                                        rcon=0.004,
                                        tolerance=45,
                                        repel=0.9,
                                        onlyCA=1)""") )

potList.append( XplorPot("BOND") )
potListIn.append( potList['BOND'] )
potList.append( XplorPot("ANGL") )
potListIn.append( potList['ANGL'] )
potList['ANGL'].setThreshold( 5 )
rampedParams.append( MultRamp(0.4,1,"potList['ANGL'].setScale(VALUE)") )
potList.append( XplorPot("IMPR") )
potListIn.append( potList['IMPR'] )
potList['IMPR'].setThreshold( 5 )
rampedParams.append( MultRamp(0.1,1,"potList['IMPR'].setScale(VALUE)") )

# Give atoms uniform weights, except for the anisotropy axis
protocol.massSetup()

# set up the term which restrains backbone angles
protocol.initDihedrals("Dih_ang_talosP")
potList.append( XplorPot('CDIH') )
potListIn.append( potList['CDIH'] )
highTempParams.append( StaticRamp("potListIn['CDIH'].setScale(10)") )
highTempParams2.append( StaticRamp("potList['CDIH'].setScale(200)") )
rampedParams.append( StaticRamp("potList['CDIH'].setScale(200)") )
potList['CDIH'].setThreshold( 5 ) #5 degrees is the default value, though
```

```
# IVM setup
# the IVM is used for performing dynamics and minimization in torsion-angle
# space, and in Cartesian space.
from ivm import IVM
dyn = IVM()

# reset ivm topology for torsion-angle dynamics
dyn.reset()

protocol.torsionTopology(dyn)

# minc used for final cartesian minimization
minc = IVM()
protocol.initMinimize(minc)
protocol.cartesianTopology(minc)

# object which performs simulated annealing
from simulationTools import AnnealIVM
init_t  = 3500.  # Need high temp and slow annealing to converge
cool = AnnealIVM(initTemp =init_t,
            finalTemp= 25,
            tempStep = 12.5,
            ivm=dyn,
            rampedParams = rampedParams)

# selection (of secondary structure elements)
# which will be used to calculate a regularized average structure
cmp_sel="""(name CA) and (resid 1:6 or resid 12:17 or resid 23:34
                or resid 38:45 or resid 48:49 or resid 57:59
                or resid 66:71  )"""

def calcOneStructure(loopInfo):
   """ this function calculates a single structure, performs analysis on the
   structure, and then writes out a pdb file, with remarks.
   """
   # randomize initial protein conformation
   from monteCarlo import randomizeTorsions
   randomizeTorsions(dyn)
   protocol.fixupCovalentGeom(maxIters=100,useVDW=1)
   protocol.initDihedrals("Dih_ang_talosP")

   # initialize parameters for high temp dynamics.
   InitialParams( rampedParams )
   # high-temp dynamics setup - only need to specify parameters which
   # differ from initial values in rampedParams
   InitialParams( highTempParams )

   # high temp dynamics
   protocol.initDynamics(dyn,
                potList=potListIn, # potential terms to use
                bathTemp=init_t,
                initVelocities=1,
                finalTime=100,     # stops at 100ps or 50000 steps
                numSteps=50000, # whichever comes first
                printInterval=100)

   dyn.setETolerance( init_t/100 )  #used to det. stepsize. default: t/1000
   dyn.run()
```

```
    InitialParams(highTempParams2)
    # high temp dynamics 2
    protocol.initDynamics(dyn,
                    potList=potList, # potential terms to use
                    bathTemp=init_t,
                    initVelocities=1,
                    finalTime=100,     # stops at 100ps or 50000 steps
                    numSteps=50000, # whichever comes first
                    printInterval=100)

    dyn.setETolerance( init_t/100 )  #used to det. stepsize. default: t/1000
    dyn.run()

    # initialize parameters for cooling loop
    InitialParams(rampedParams)

    # initialize integrator for simulated annealing
    protocol.initDynamics(dyn,
                    potList=potList,
                    numSteps=3000,    #at each temp: 100 steps or
                    finalTime=0.5 ,      # .5ps, whichever is less
                    printInterval=100)

    # perform simulated annealing
    cool.run()

    # final torsion angle minimization
    protocol.initMinimize(dyn, printInterval=50)
    dyn.run()

    # final all-atom minimization
    protocol.initMinimize(minc, potList=potList, dEPred=10)
    minc.run()

    p_tess.forceTessellation()
    #do analysis and write structure
    loopInfo.writeStructure(potList)
    pass

from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
            pdbTemplate=outFilename,
            structLoopAction=calcOneStructure,
            genViolationStats=1,
            averagePotList=potList,
            averageSortPots=potList,
            averageTopFraction=0.1, # report only on best 10% of structures
            averageContext=FinalParams(rampedParams),
            averageFilename="SCRIPT_ave.pdb", # generate regularized ave structure
            averageFitSel=cmp_sel,
            averageCompSel=cmp_sel).run()
```

Example of relaxation restraints files
(a) T1 data:

```
1  2 GLN N 0.477
2  3 ILE N 0.453
3  4 PHE N 0.439
4  5 VAL N 0.471
5  6 LYS N 0.462
6  7 THR N 0.460
7  8 LEU N 0.476
8  9 THR N 0.510
9 10 GLY N 0.495
```

(b) T2 data:

```
1  2 GLN N 0.167
2  3 ILE N 0.162
3  4 PHE N 0.158
4  5 VAL N 0.175
5  6 LYS N 0.170
6  7 THR N 0.167
7  8 LEU N 0.196
8  9 THR N 0.195
9 10 GLY N 0.203
```

Table S2 Breakdown of distance restraints

Structure	Hydrogen bonds[a]		NOEs involving NH and methyl groups			
	$\|i - j\| = 4$	$\|i - j > 4$	intraresidue	interresidue		
				$\|i - j\| = 1$	$1 < \|i - j\| \leq 5$	$\|i - j\| > 5$
GB3	12	23	-	-	-	-
Ubiquitin	10	18	-	-	-	-
EIN	87	27	69	312	278	145

[a]There are two distance restraints for each hydrogen bond: $r_{N\text{-}O}$ = 2.4-3.5 Å and $r_{HN\text{-}O}$ = 1.5-2.8 Å.
[b]The NOE derived distance restraints are classified into four distance ranges, 1.8-2.9, 1.8-3.5, 1.8-5.0 and 1.8-6.0 Å, corresponding to strong, medium, weak and very weak NOEs with an additional 0.5 Å per methyl group added to the upper bounds to account for the sharper methyl group resonances.

Table S3 Diffusion tensor parameters.

Structure	T_{diff}^{app}	Rotational correlation Time (ns)	Anisotropy	Rhombicity
GB3 (nominal experimental temperature 297 K)				
Reference[a]	N/A	3.35	1.30	0.15
Ten lowest energy structures[b]				
#50	301.21	3.38	1.27	0.12
#51	300.11	3.40	1.26	0.06
#23	300.26	3.36	1.25	0.08
#95	300.02	3.38	1.25	0.17
#99	300.65	3.30	1.26	0.13
#93	300.21	3.37	1.25	0.06
#20	300.36	3.35	1.24	0.11
#54	301.69	3.33	1.27	0.07
#13	300.24	3.33	1.22	0.18
#6	300.53	3.33	1.24	0.16
Average	300.53±0.53	3.35±0.03	1.25±0.02	0.11±0.05
Ubiquitin (nominal experimental temperature 300 K)				
Reference[a]	N/A	4.14	1.19	0.18
Ten lowest energy structures[b]				
#10	304.26	4.07	1.14	0.46
#46	304.13	4.07	1.10	0.20
#96	304.39	4.06	1.12	0.37
#56	304.01	4.05	1.08	0.51
#13	303.76	4.01	1.08	0.26
#58	304.67	4.06	1.09	0.87
#57	303.68	4.10	1.12	0.51
#95	303.47	4.07	1.15	0.51
#35	303.96	4.05	1.15	0.22
#70	303.94	4.08	1.13	0.34
Average	304.03±0.35	4.06±0.02	1.12±0.03	0.43±0.20
EIN (nominal experimental temperature 316 K)				
Reference[a]	N/A	10.65	1.67	0.06
Ten lowest energy structures[b]				
#92	316.09	10.45	1.74	0.02
#43	317.04	10.63	1.76	0.01
#61	318.16	10.48	1.72	0.02
#74	317.01	10.51	1.77	0.01
#58	317.82	10.55	1.75	0.01
#44	316.77	10.61	1.73	0.02
#24	317.86	10.43	1.77	0.01
#86	318.09	10.55	1.80	0.01
#23	317.04	10.46	1.70	0.01
#72	318.35	10.46	1.77	0.01
Average	317.42±0.74	10.51±0.07	1.75±0.03	0.01±0.01

[a]The diffusion tensor parameters for the references X-ray structures of GB3 (1IGD), ubiquitin (1UBQ) and EIN (1ZIM) were calculated by best-fitting the experimental ^{15}N-R_2/R_1 relaxation data to the N-H bond vector

Footnotes to Table S3 (cont.)

orientations (with amide protons added to the coordinates in standard geometry). Outliers with relative deviations larger than 1.5σ between observed and calculated ^{15}N-R_2/R_1 ratios (corresponding to 13% of the data) were excluded from the fits as described previously (Ryabov, Y.; Clore, G.M.; Schwieters, C.D. *J. Am. Chem. Soc.* **2010**, *132*, 5987-5989). The excluded residues are listed in Table S4.

[br]The diffusion tensor parameters for the 10 lowest energy structures were derived from the molecular shape and size of the structure obtained by representing the surface of the protein by an equivalent ellipsoid and then applying Perrin's equations to calculate the diffusion tensor from the dimensions and orientation of the ellipsoid. (Ryabov, Y.; Suh, Y.-J.; Grishaev, A.; Clore, G.M.; Schwieters, C.D. *J. Am. Chem. Soc.* **2009**, **131**, 9522-9531). The apparent diffusion tensor temperature, T_{diff}^{app}, is a parameter that is optimized during the course of simulated annaling and collects uncertainties in sample temperature, viscosity and protein hydration layer description. Outlier ^{15}N-R_2/R_1 ratios are excluded iteratively during the course of the structure determination calculations as described in the main text. The final list of excluded residues is provided in Table S4.

Table S4. Statistics of excluded residues with outlying ^{15}N-R_1/R_2 ratios[a]

Structure	Excluded residues
GB3	
Reference X-ray structure (1IGD)	12, 20, 31, 39, 41, 46, 52
Ten Lowest energy structures	
#50	20, 31, 36, 39
#51	29, 31, 32, 36, 39
#23	31, 36, 39, 41
#95	22, 29, 31, 32, 36, 39, 41
#99	31, 36, 39, 41
#93	31, 39, 41
#20	22, 29, 31, 32, 36, 39
#54	29, 31, 32, 36, 39, 41
#13	31, 36, 39
#6	22, 29, 31, 32, 36, 39
Ubiquitin	
Reference X-ray structure (1UBQ)	8, 10, 11, 23, 25, 73-76
Ten lowest energy structures	
#10	25, 73-76
#46	25, 73-76
#96	25, 73-76
#56	25, 73-76
#13	25, 73-76
#58	25, 73-76
#57	25, 73-76
#95	25, 73-76
#35	25, 73-76
#70	25, 73-76
EIN	
Reference X-ray structure (1ZYM)	5, 36, 70, 77, 82, 89, 92, 116, 135, 158, 175, 184, 197, 203, 206, 246
Ten lowest energy structures	
#92	57, 77, 158, 204, 206
#43	4, 6, 77, 175, 182, 204
#61	4, 7, 77, 184, 204
#74	77, 82, 203, 230
#58	77, 204, 207, 223, 230
#44	36, 39, 77, 204
#24	4, 6, 77, 184, 204, 230
#86	77, 204, 214, 230
#23	4, 7, 77, 203, 213
#72	5, 70, 82, 204, 207

[a]Outliers for the known reference structures are excluded as described previously (Ryabov, Y.; Clore, G.M.; Schwieters, C.D. *J. Am. Chem. Soc.* **2010**, *132*, 5987-5989). See footnote *a* of Table S3. Outliers for the ten lowest energy structures are excluded iteratively during the course of structure determination as described in the main text. Note that slightly fewer residues are excluded by the latter than the former, as small reorientations of N-H bond vectors (while maintaining idealized covalent geometry) during the course of simulated annealing allows for improved agreement between calculated structures and the input relaxation data.

Footnotes to Table S4 (cont.)

[b]The secondary structure elements for GB3 are as follows. β-strands: 1-8, 13-20, 42-46 and 51-53; α-helix 23-36.

[c]The secondary structure elements for ubiquitin areas follows. β–strands: 1-7, 12-17, 38-45 (includes a turn), 48-49 and 66-71; α-helix 23-34; turn: 57-60. Note residues 73-76 comprise a disordered C-terminal tail.

[d]The secondary structure elements for EIN are as follows. β-strands: 12-18, 156-159, 173-180 (includes a turn), 201-202, 217-220 and 27-229; α-helix: 36-65, 67-80, 83-94, 100-112, 121-141, 149-151, 165-170, 191-197, 208-211, 223-239 and 244-247.