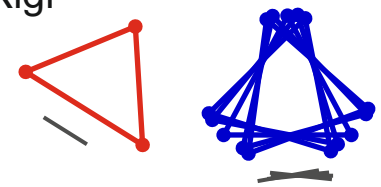
MMMx: RigiFlex modelling

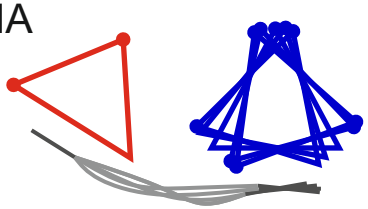
G. Jeschke

ETH Zürich, Dep. Chemistry & Applied Biosciences

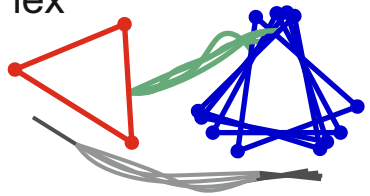
Rigi



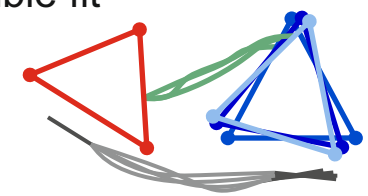
FlexRNA

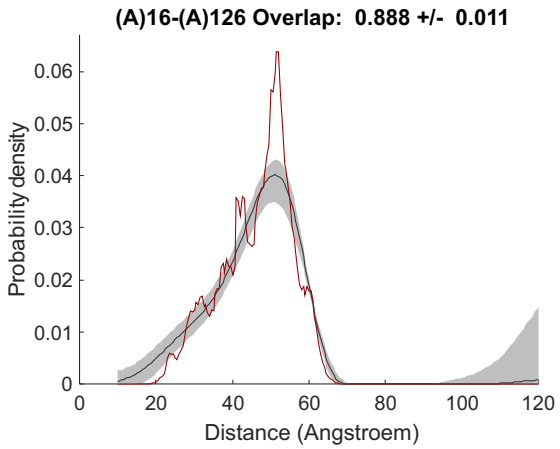


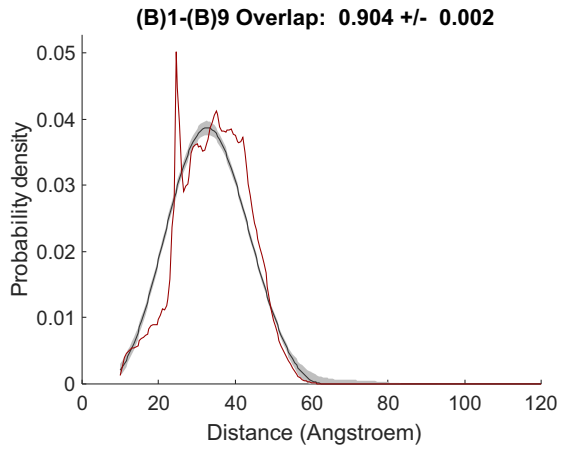
Flex

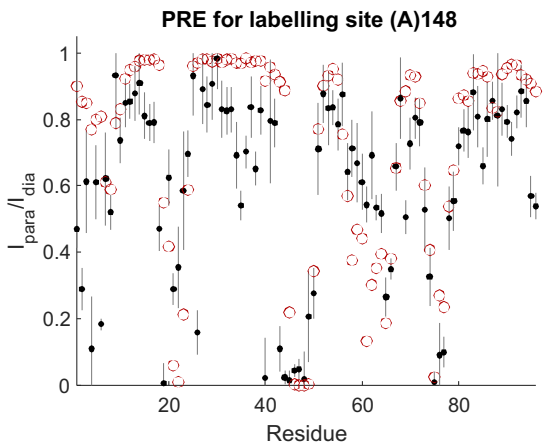


Ensemble fit









Modelling SRSF1 in complex with UCAUUGGAU

Steps

- run `Rigi` module with hierarchical clustering to generate an ensemble of rigid-body arrangements
- run `FlexRNA` to generate a two-nucleotide linker UU between binding motifs CA and GGA
- run `FlexRNA` to generate single 5'-terminal nucleotide U
- run `FlexRNA` to generate single 3'-terminal nucleotide U
- run `Flex` to generate peptide linker from residue 90 to 120
- refine all conformers by `Yasara`

At this point, we have generated a *raw ensemble* by RigiFlex

- run `EnsembleFit` with the raw ensemble, DEER distance distribution restraints, and PRE restraints

Dependences on third-party software

- `SCWRL4` for generating amino acid side groups in `Flex`
- `Yasara` for refining conformers (proprietary software)
- refinement by `GROMACS` (free) is possible, but requires manual force-field choice

Making rigid-body arrangements

```
!rigi % call the Rigi module
  rbtemplate BSRS.pdb % load the rigid-body template
  separate on % separate the rigid bodies from each other for spin labelling
  maxtrials 10000 % make 10000 trials within the distance distributions
  models 200 % generate up to 200 models
  save SRSF1_rba % save in MMMx internal rigid-body arrangement format with file name SRSF1_rba.mat
  rigid (A) (B) % define the first rigid body by chains A and B of the template
    (A)16 mtsl % first reference point of first rigid body, labelled by MTSL
    (A)37 mtsl % second reference point of first rigid body, labelled by MTSL
    (A)72 mtsl % third reference point of first rigid body, labelled by MTSL
  .rigid % close block key
  rigid (C) (D) % define the first rigid body by chains C and D of the template
    (D)126 mtsl % first reference point of second rigid body, labelled by MTSL
    (D)148 mtsl % second reference point of second rigid body, labelled by MTSL
    (D)169 mtsl % third reference point of second rigid body, labelled by MTSL
  .rigid % close block key
  ddr mtsl % define 7 core restraints between reference points in different rigid bodies
    (A)16     (D)148     32.5     06.0 % Gaussian restraint
    ...  % there are six more lines as the previous one
  .ddr % close block key
```

Further specifications for rigid-body arrangements

```
  plink % specify the length of a peptide linker
    (A)89  (D)121 32  % 32 residues anchored at sites (A)89 and (D)121, <=3.8 Å/residue
  .plink

  nlink % specify the length of a nucleic acid linker
    (B)3   (C)6  3  16 % 3 nucleotides anchored at sites (B)3 and (C)6, maximum length 16 Å
  .nlink

  superimpose 2 % superposition is onto rigid body 2
.rigi % close module
```

- Laura Esteban Hofer was running 50000 trials on Euler and she requested 6000 models
- when running `Flex` on Euler with a fresh MMMx installation, specify the path to SCWRL4 in the `Flex` block by
  `scwrl4` *pathname*

  (this is generally necessary on Linux or Mac, Matlab fails to find the correct path)

Completing the RNA

```
!flexrna 0.75 1 0.016667 % 75% coverage of distributions, 1 model, maximum of 1 min per conformer
  expand SRSF1_rba % the input conformers are generated by expanding rigid-body arrangements
  sequence 4 5 UU % add nucleotides number 4 and 5 with sequence UU
  save SRSF1_short_UU_all % save the models with basis file name 'SRSF1_short_UU_all'
  anchor_5p (B)3 % the 5'-terminal anchor nucleotide is nucleotide 3 in chain B
  anchor_3p (C)6 % the 3'-terminal anchor nucleotide is nucleotide 6 in chain C
.flexrna % close module

!flexrna 0.75 1 0.025 % 75% coverage of distributions, 1 model, maximum of 1.5 min per conformer
  addpdb SRSF1_short_UU_all*.pdb  % load conformers generated by the previous module call
  sequence 1 1 U % segment is a single nucleotide U with residue number 1
  save SRSF1_short_U1 % save the models with basis file name 'SRSF1_short_U1'
  anchor_3p (B)2 % the 3'-terminal anchor nucleotide is nucleotide 2 in chain C

  ddr dota-gd r5p % specify distance distribution restraints, dota-gd label on protein, r5p on RNA
    (A)16   (B)1  34.9  04.1 % Gaussian restraint to RRM1
    (D)148  (B)1  26.5  08.5 % Gaussian restraint to RRM2
  .ddr % close block key
.flexrna % close module

% there is one more FlexRNA block for adding the 3'-terminal nucleotide
```

Adding the peptide linker

```
!flex 0.75 1 0.25 % 75% coverage of distributions, 1 model, maximum of 15 min per conformer
  addpdb SRSF1_short_U9*.pdb % load conformers from previous section
  sequence 90 120 RSGRGTGRGGGGGGGGGAPRGRYGPPSRRSE % specify residue numbers and sequence
  n_anchor (A)89 % the N-terminal anchor is residue 89 in chain A
  c_anchor (D)121 % the C-terminal anchor is residue 121 in chain D
  save SRSF1_short_RNA SRSF % save conformers with basis file name 'SRSF1_short_RNA'

  ddr mtsl % specify peptide-peptide distance distribution restraints (MTSL label pairs)
    (A)16  107     29.3  08.7  @deer\C16_A107C_short_med_distr.dat % site 107 is newly generated
    107    (D)148  25.0  07.3  @deer\A107C_C148_short_med_distr.dat % full distribution is specified
    (A)37  107     39.7  08.0  @deer\Y37C_A107C_short_med_distr.dat
  .ddr   % close block key

  ddr dota-gd r5p % specify peptide-RNA distance distribution restraints between dota-gd and r5p
    107  (B)1  27.2   10.8  @deer\A107C_U1_short_med_distr.dat
  .ddr

  ddr dota-gd r3p % specify peptide-RNA distance distribution restraints between dota-gd and r3p
    107   (B)9  31.7  11.8  @deer\A107C_U9_short_med_distr.dat
  .ddr
.flex % close module
```

Refining conformer models

```
!yasara 1 % allow for up to 1 hour for refinement
   addpdb SRSF1_short_RNA_i*_m1.pdb % process all output conformers from the previous section
   save SRSF1_short_refined % save output with basis file name 'SRSF1_short_refined'
.yasara % close module

# report % open log file in editor
```

- Yasara cannot be stopped by MMMx
- if Yasara runs take longer than the specified maximum time, more and more Yasara instances are generated
- this can severely slow down a desktop computer
- avoid the `# report` statement if you run on a server (Euler), use `console` mode

Fitting populations and contracting the ensemble

```
!ensemblefit
   addpdb SRSF1_short_refined_m*.pdb % use all refined conformers in the raw ensemble
   interactive % display figure that visualizes fit progress
   plot % plot figures on fit quality
   csv % save fits in comma-separated value files
   save SRSF1_UCAUUGGAU_ensemble_fit.ens % output name for MMMx ensemble list
   ddr mtsl % specify distance distribution restraints protein to protein sites
        (A)16     (A)148     32.5     06.0     @deer\C16_C148_short_med_distr.dat
        ... % more lines as the previous one
   .ddr
   ddr dota-gd r5p % specify distance distribution restraints protein site to 5'-terminus of RNA
       (A)107     (B)1     27.2     10.8     @deer\A107C_U1_short_med_distr.dat
       ... % more such lines
    .ddr
   ddr dota-gd r3p % specify distance distribution restraints protein site to 3'-terminus of RNA
       (A)107     (B)9     31.7     11.8     @deer\A107C_U9_short_med_distr.dat
       ...  % more such lines
  .ddr
  ddr r5p r3p % specify distance distribution restraint between RNA termini
       (B)1     (B)9  32.6     10.5     @deer\1p2WT_U1_U9_short_med_distr.dat
  .ddr
```

Specifying PRE restraints

```
   % PRE ratio data C16
   % pre label   site  larmor td     R2dia taui taur  maxrate
   pre   mtsl    (A)16 600.13 12.812 49.66 0.50 11.15 170
           (A)18   0.941   0.280 % proton site, PRE ratio, standard deviation
           (A)19   0.529   0.134 % proton site, PRE ratio, standard deviation
           ...
   .pre % close PRE block
   % PRE ratio data C148
   % pre label   site   Larmor td     R2dia taui taur  maxrate
   pre   mtsl    (A)148 600.13 12.812 49.35 0.50 11.15 170
           (A)17   0.470   0.057 % proton site, PRE ratio, standard deviation
           (A)19   0.289   0.064 % proton site, PRE ratio, standard deviation
           ...
   .pre % close PRE block
.ensemblefit % close module
```

- the PRE block specifies the label type, spin label site, proton Larmor frequency in MHz, total INEPT delay td in ms, relaxation rate R2dia for the diamagnetic sample in s^{-1}, correlation time taui of internal label motion in ns, rotational correlation time taur of the protein in ns, and a maximum relaxation rate enhancement maxrate in s^{-1}