Rigid Body Modeling : Overview and SASREF

Srinivas Chakravarthy
IIT/CSRRI
Staff Scientist, BioCAT
Sector 18, Advanced Photon Source

The Need for Rigid Body Modeling

- Biological macromolecules in their physiologically relevant form - multi-component complexes or large multi-domain proteins.
- Large complexes often not feasible to study using high-resolution techniques (crystallography, NMR etc…)
- High Resolution Structures of individual sub-units or individual domains from large multi-domain proteins can be docked into low-resolution envelopes (either EM or SAXS) of the entire complex or full-length protein.
- Advantage with SAXS over EM - No size restrictions.

Rigid-Body Modeling Using SAXS - Evolution

- Recent (~ the past decade) developments in SAXS data analysis enable *ab initio* calculation of overall macromolecular shape.
- EM – High resolution structures of subunits docked into envelope.
- SAXS – Spatial arrangement of subunits is reached at by direct fitting of scattering data from the complex.
- Wall et al., (2000) – Used ellipsoids to represent subunits and to find the relative arrangement of these subunits.
- Constrained Fit Procedure – Screening of thousands of bead models along with AUC results.

Rigid Body Modeling - Theory

$$I(s) = \left\langle \left| \sum_{k=1}^{K} A^{(k)}(s) \right|^2 \right\rangle_{\Omega},$$

$$A^{(k)}(s) = \exp(isr_k)\Pi(\alpha_k\beta_k\gamma_k)[C^{(k)}(s)]$$

Complex with K subunits with known structure, scattering amplitude from each subunit denoted as $C^{(k)}(s)$, $s = 4\pi\sin(\theta) / \lambda$. Subunits rotated and translated as rigid bodies changing their scattering amplitudes. I(s) scattering intensity of the entire complex. $A^{(k)}(s)$ denotes scattering amplitude of the kth rigid body at the given position, $\langle\ldots\rangle_{\Omega}$ - spherical avg. in reciprocal space. $\Pi(\alpha_k\beta_k\gamma_k)$ is the rotational operator. Modified scattering $A^{(k)}(s)$ depend on rk (vector of the shift) and the Euler rotation angles – α_k β_k and γ_k.

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left| \sum_{k=1}^{K} A_{lm}^{(k)}(s) \right|^2$$

$A_{lm}^{(k)}(s)$ – partial scattering amplitudes of the k^{th} rigid body, depend on $C_{lm}^{(k)}(s)$ which is the scattering amplitudes in reference position and orientation and can be calculated by CRYSOL.

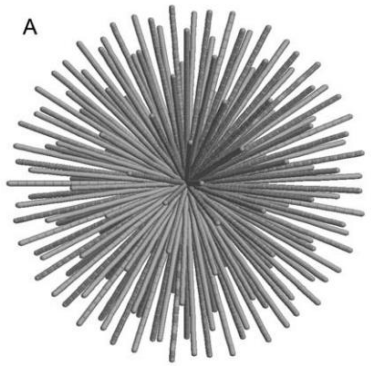
$$\chi^2 = \frac{1}{N-1} \sum_{j} \left[\frac{I_{exp}(s_j) - cI(s_j)}{\sigma(s_j)} \right]^2$$

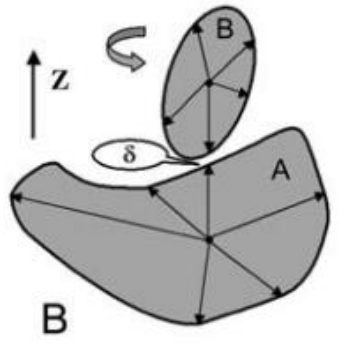
The goodness of fit provided by a given arrangement of bodies to the experimental data $I_{exp}(s)$ is measured by the discrepancy depicted in this equation. N is the number of experimental points, c is a scaling factor, and $\sigma(s_j)$ Is the experimental error at the momentum transfer s_j.

The target function $E = \chi^2 + \Sigma\alpha_i P_i$ used to generate physically sound models, $\alpha_i P_i$ – penalty terms that formulate the requirements of interconnectivity and absence of overlaps and also incorporate outside information such as inter-residue distances.

Homo- and Hetero- Dimers

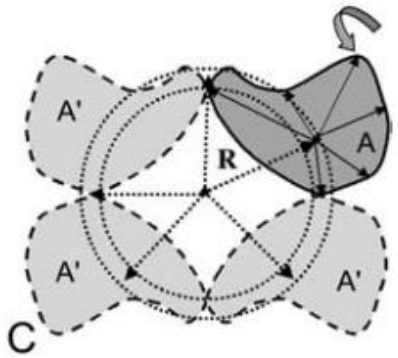
- Shapes of two monomers represented by angular envelope functions $F(\omega)$ where ω is the solid angle in real space. Envelopes generated by CRYSOL.
- A sequence of Fibonacci numbers for the values of polar angles defining sampling directions.
- Both monomers centered at the origin, the first monomer is rotated to bring the j^{th} direction of Fibonacci grid ω_j to the z axis, and the other rotated to bring the i^{th} direction of its Fibonacci grid ω_i antiparallel to the z axis. [Homodimers – only j = i case is taken to ensure the symmetric arrangement of monomers].
- Second monomer is shifted along the z axis by $F^{(1)}(\omega_j) + F^{(2)}(\omega_i) + \delta$ and rotated about this axis by an angle $0 < \psi < 2\pi$ with a discrete angular step. The offset $\delta \sim 0.3$nm – ensures reasonable contact between monomer surfaces and avoids steric clashes.
- Implemented in DIMFOM.





Symmetric Oligomers

- Appropriate positioning of monomer and generated symmetry mates – 6 parameters – shift vector r = (r, ψ, and ϕ) and rotation angles (α, β, and γ). Scattering intensity calculations accelerated due to symmetry.
- Magnitude of the shift vector r is constrained by the experimental value of the Rg of the entire oligomer R_g^{exp}. Average value of the shift from origin is $<|r_0|> = \sqrt{(R_g^{exp})^2 + (R_g^{mon})^2}$, where R_g^{mon} is the radius of gyration of the monomer.
- Information about distances between specific residues in the oligomer can be used as a restraint.
- Minimized Target function $E = \chi^2 + \alpha_{cross}P_{cross} + \alpha_{cont}P_{cont}$, P_{cross} – prevents inter-subunit overlap, P_{cont} – inter-residue distances.



Implemented in GLOBSYMM

Quaternary Structure of Multi-subunit Complexes - SASREF

- Simulated Annealing performs random modifications of the system by moving to configurations with a lower scoring function E.
- Beginning – high temperature – moves almost random, and towards the end a configuration with the minimum E is reached.
- $E = \Sigma(\chi^2)_i + \alpha_{dis}P_{dis} + \alpha_{cross}P_{cross} + \alpha_{cont}P_{cont,}$
- P_{dis} – ensures interconnectivity of the model, P_{cross} – absence of overlaps between subunits, P_{cont} – information about distances between residues or nucleotides.
- Sequence elements can be specified that are expected to be involved in inter-subunit contacts.
- For symmetric particles with K subunits in the asymmetric unit, appropriate symmetry mates are generated to build the model of the entire complex. Only the symmetry independent part is modified during simulated annealing.

Rigid Body Modeling with Addition of Missing Fragments: BUNCH

- High-resolution models often have missing fragments both in multisubunit assemblies and multidomain proteins consisting of globular domains linked by flexible loops.
- Aim is to simultaneously find best relative orientation of the subunits/domains and the best possible conformation of the flexible linkers.
- The interconnected chains are composed of dummy residues. A loop or protein fragment with unknown structure substituted by a flexible chain of interconnected residues with spacing 0.38nm.
- Similar to SASREF, simulated annealing is used for global minimization of the target function –

$E = \Sigma(\chi^2)_i + \alpha_{cross}P_{cross} + \alpha_{ang}P_{ang} + \alpha_{dih}P_{dih} + \alpha_{ext}P_{ext}$, P_{cross} – avoids inter-domain overlaps, P_{ang} and P_{dih} ensure the proper bond and dihedral angles in the dummy residue chains. P_{ext} is introduced to avoid over-extended conformations of the dummy residue loops.

- In complexes, missing loops AND shapes of missing subunits can be reconstructed.
- Warning – Loop configuration is usually an average conformation of flexible loops and must therefore not be considered their actual structure.

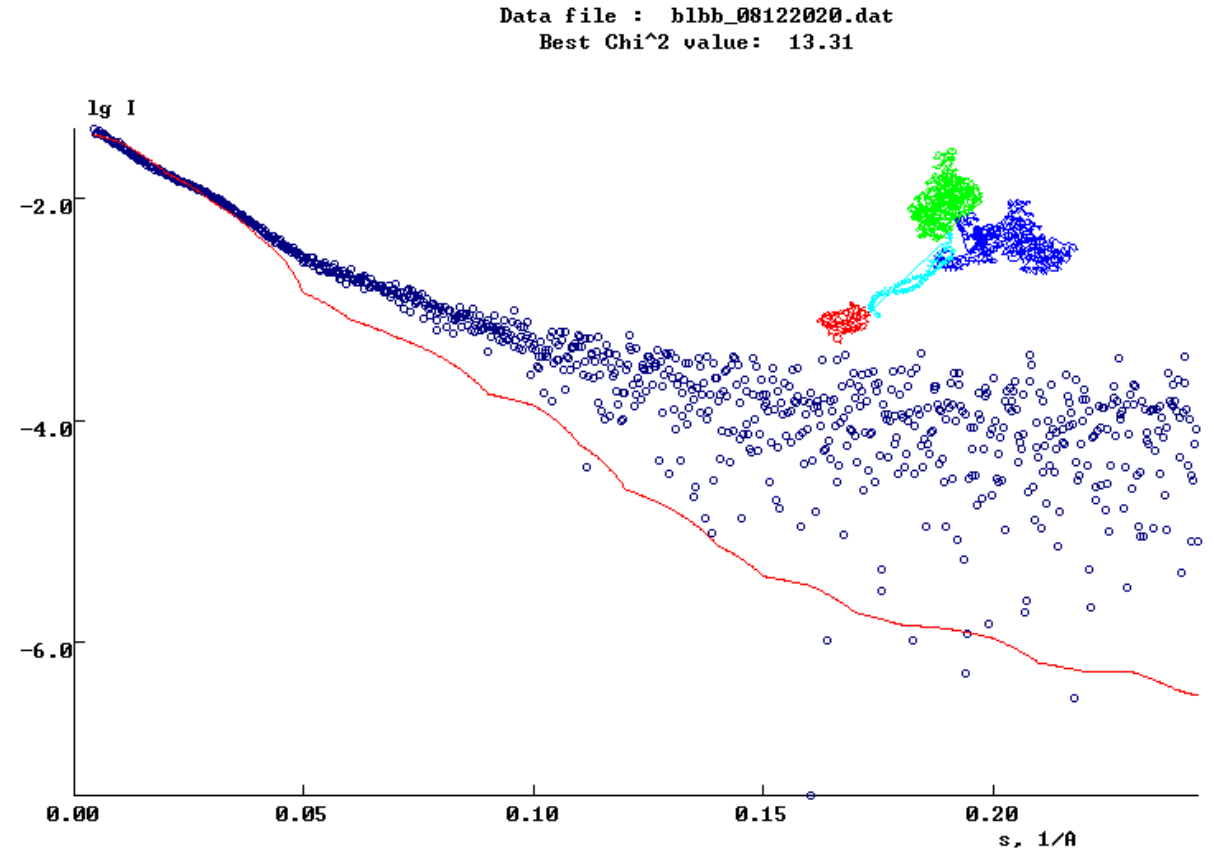
Running SASREF – Interactive Mode

Screen Text	Default	Asked in USER-mode?	Description
Computation mode (User or Expert)	USER	Y	Mode selection.
Log file name	N/A	Y	Project identifier, will be used as a prefix for all output file names.
Enter project description	N/A	Y	Any text that will be stored in the log file.
Input total number of subunits	1	Y	Number of separate rigid bodies in the asymmetric part of the complex.
Symmetry: P1...19 or Pn2 (n=1,..,12)	P1	Y	Supported symmetries are: P1, P2-P19 (nineteen-fold), P222, P32-P(12)2, as well as cubic types (P23, P432) and icosahedral symmetry (Pico). The n-fold axis is typically Z, if there is in addition a two-fold axis, it coincides with Y.
Input total number of scattering curves	1	Y	If in addition to the whole complex, some sub-complexes (and their scattering profiles) are available, they can be fitted simultaneously assuming the same arrangement of subunits in all the constructs.
Use Kratky Geometry	N	N	If the answer is Yes, the computed curves will be smeared to fit the data from Kratky camera.
Input first & last subunits in 1-st construct	1,var	Y	The range of subunits present in the given construct (one scattering curve=one construct; one subunit=one rigid body=one pdb file). This question is asked for each construct, i.e. the number of times equals to the total number of scattering curves (answer to the previous question). The default answer is from 1 to the total number of scattering curves.
Enter file name, 1-st experimental data <dat >	N/A	Y	The name of the data file containing the experimental SAXS profile of a certain construct. The question is asked for each construct.
Angular units in the input file : 4*pi*sin(theta)/lambda [1/angstrom] (1) 4*pi*sin(theta)/lambda [1/nm] (2) 2* sin(theta)/lambda [1/angstrom] (3) 2* sin(theta)/lambda [1/nm] (4)	1	Y	Formula for the scattering vector in the data file and its units. The question is asked for each construct.
Fitting range in fractions of Smax	1.0	Y	Percentage of the scattering curve to fit, starting at the first point. Default is the entire curve. The question is asked for each construct.
Amplitudes, 1-st subunit <alm>	N/A	Y	The name of the file with partial scattering amplitudes of a certain subunit computed by CRYSOL. This question is asked for each subunit, i.e. the number of times equals to the total number of subunits.
Initial rotation by alpha	0.0	Y	The user can specify an arbitrary initial rotation by Euler angle Alpha. By default, no rotation is made, i.e. the reference orientation in the PDB file is used as a starting one. This question is asked for each subunit.
Initial rotation by beta	0.0	Y	The user can specify an arbitrary initial rotation by Euler angle Beta. This question is asked for each subunit.
Initial rotation by gamma	0.0	Y	The user can specify an arbitrary initial rotation by Euler angle Gamma. This question is asked for each subunit.

Initial shift along X	var	Y	The user can specify an arbitrary initial shift along the X-axis of the orthogonal coordinate system. By default, the subunit is shifted to the position as it appears in the PDB file. Another reasonable option is to place the subunit at the origin (`0.0 0.0 0.0`) and let the program build the complex "from scratch". This question is asked for each subunit.
Initial shift along Y	var	Y	The user can specify an arbitrary initial shift along the Y-axis of the orthogonal coordinate system. This question is asked for each subunit.
Initial shift along Z	var	Y	The user can specify an arbitrary initial shift along the Z-axis of the orthogonal coordinate system. This question is asked for each subunit.
Movements limitations of subunit: N/F/X/Y/Z/D?	N	Y	It is possible to fix the subunit in the original position/orientations (`F`), e.g. to keep the desired mutual arrangement between the certain subunits, or to move/rotate the subuntis only along specified axes: `X`, `Y`, `Z` or the cube's diagonal (`D`). If the answer is `N`, no restrictions are applied. This question is asked for each subunit.
Spatial step in angstroems	5.0	N	Maximal random shift of a subunit at a single modification of the system in the course of simulated annealing. This question is asked for each subunit.
Angular step in degrees	20.0	N	Maximal random rotation angle of a subunit at a single modification of the system in the course of simulated annealing. Setting it to zero may be useful to keep the mutual orientations of certain subunits, e.g. if NMR RDC data are available. This question is asked for each subunit.
Disconnectivity penalty weight	10.0	N	How much the Disconnectivity Penalty shall influence the acceptance or rejection of a mutation. A value of `0.0` disables the penalty. If unsure, use the default value. If not interconnected arrangement of the subunits is observed, try increasing this penalty weight.
File name, contacts conditions, CR for none <.cnd >	empty	Y	If the information on interface between certain subunits in terms of contacting residues is available, it may be used as a modeling restraint. The information is provided in a file with special format. By default no information is given.
Contacts penalty weight	10.0	N	How much improper contacts shall influence the acceptance or rejection of a mutation. If unsure, use the default value. If desired interfaces are not obtained, try increasing this penalty weight. This question is only asked if the contacts conditions file is provided.
Expected particle shape: Prolate, Oblate, or Unknown	UNKNOWN	Y	If, due to prior studies, it is known that the particle's shape shall be either `PROLATE` or `OBLATE`, one may use the anisometry option to enforce a penalty on particles that do not correspond with the expected anisometry. By default, anisometry is 'UNKNOWN'.
Anisometry penalty weight	1.0	N	How much improper anisometry shall influence the acceptance or rejection of a mutation. If unsure, use the default value. This question is skipped if the Expected particle shape is 'UNKNOWN'.
Expected direction of anisometry: aLong Z, aCross Z, or Unknown	UNKNOWN	Y	This question is only asked if the Expected particle shape is not 'UNKNOWN' and the symmetry is 'P2'. The user can specify if the symmetry axis coincides with (`ALONG`) or perpendicular to (`ACROSS`) the anisometry axis.

Shift penalty weight	1.0	N	How much shift from the origin of the entire complex shall influence the acceptance or rejection of a mutation. A value of 0.0 disables the penalty. If unsure, use the default value. This penalty is necessary to keep the model close to the origin so that the higher order harmonics are not lost and the scattering is computed accurately.
Initial annealing temperature	10.0	N	Starting temperature of simulated annealing protocol.
Annealing schedule factor	0.9	N	Factor by which the temperature is decreased; 0.9 is a good average value. If slower cooling is wanted increase the value (e.g. to 0.95).
Max # of iterations at each T	var	N	Finalize temperature step and cool after this many iterations at the latest. The default value is 5000* total number of subunits.
Max # of successes at each T	var	N	Finalize temperature step and cool after at most this many successful mutations. The default value is 500* total number of subunits.
Min # of successes to continue	var	N	Stop simulated annealing if not at least this many successful mutations within a single temperature step can be done. The default value is 50* total number of subunits.
Max # of annealing steps	100	N	Stop if simulated annealing is not finished after this many steps. The slower the system is cooled, the more temperature steps are required.

Graphics Window



Command Window

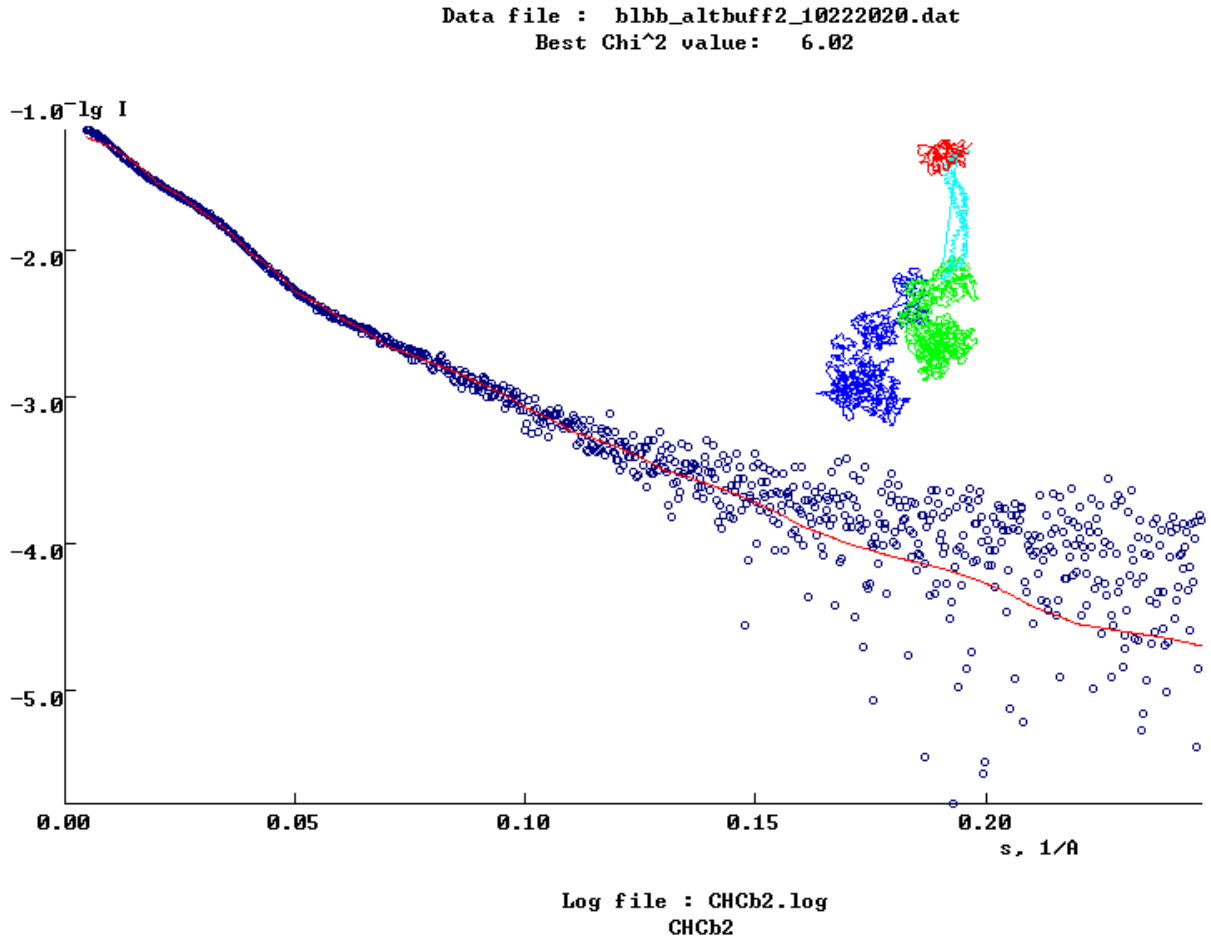
```
 Initial shift along Z ..................... <        3.277 >:
 Movements limitations of subunit: N/F/X/Y/Z/D? <            N >:
 Spatial step in angstroems ................ <        5.000 >:
 Angular step in degrees ................... <       20.000 >:
 Cross value ..................................................... : 2.304e-3
 Discontiguity value ............................................. : 0.2877
 Cross penalty weight ...................... <       10.000 >:
 Disconnectivity penalty weight ......... <       10.000 >:

          ***  PLEASE SELECT THE CONTACTS CONDITIONS FILE  ***

 Working directory: C:\Users\schakrav\Documents\20200330MyokardiaSAXSdata\CHC_pdbs
\
 File to be opened: contacts.cnd
Condition #  1: Distance  30.000
   Between subunit #  1, Residues from  CA  GLY A 823  to  CA  GLY A 823
       and subunit #  3, Residues from  CA  LYS C 841  to  CA  LYS C 841
Condition #  2: Distance  20.000
   Between subunit #  2, Residues from  CA  LEU B 840  to  CA  LEU B 840
       and subunit #  3, Residues from  CA  GLU C 844  to  CA  GLU C 844
Condition #  3: Distance   5.000
   Between subunit #  3, Residues from  CA  LEU C 961  to  CA  LEU C 961
       and subunit #  4, Residues from  CA  SER D   2  to  CA  SER D   2
 Contacts conditions penalty ..................... : 181.3
 Contacts penalty weight ................... <       10.00 >:
 Expected particle shape: <P>rolate, <O>blate,
  or <U>nknown ........................... <      Unknown >:
 Shift penalty is normalized by ................... : 44.62
 Shift penalty ..................................... : 8.237
 Shift penalty weight ...................... <       1.000 >:
 Total penalty ..................................... : 1824.
 1-st curve:
 Theoretical  points from           1  to          26  used
 The best chi^2 values:29.50164
 Initial fVal ...................................... : 1853.
 Initial annealing temperature ......... <       10.00 >:
 Annealing schedule factor ............. <      0.9000 >:
 Max # of iterations at each T ......... <       20000 >:
 Max # of successes at each T .......... <        2000 >:
 Min # of successes to continue ........ <         100 >:
 Max # of annealing steps ............... <         100 >:
  ====  Simulated annealing procedure started  ====
 j:   1 T: 0.100E+02 Suc:  2000 Eva:     3310 CPU:  0.629E+02 F:21.5265 Pen:  8.22
05
 The best chi^2 values:13.30593
```

ile Edit View Window Help

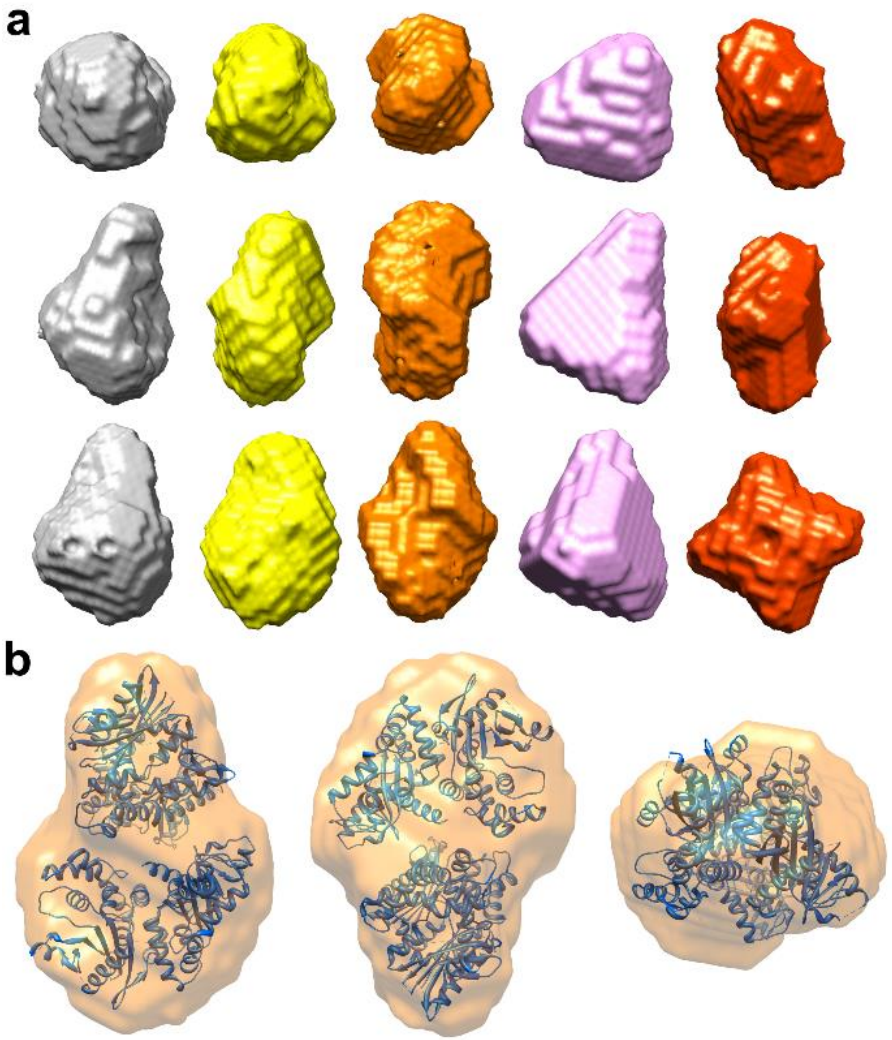
Graphics Window



Command Window

```
 j:  51 T: 0.515E-01 Suc:  1104 Eva:   565991 CPU:  0.110E+05 F: 6.4039 Pen:  0.05
80
 The best chi^2 values: 6.34589
 j:  52 T: 0.464E-01 Suc:  1113 Eva:   585991 CPU:  0.114E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  53 T: 0.417E-01 Suc:  1156 Eva:   605991 CPU:  0.118E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  54 T: 0.376E-01 Suc:  1041 Eva:   625991 CPU:  0.122E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  55 T: 0.338E-01 Suc:   925 Eva:   645991 CPU:  0.126E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  56 T: 0.304E-01 Suc:  1045 Eva:   665991 CPU:  0.130E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  57 T: 0.274E-01 Suc:   850 Eva:   685991 CPU:  0.134E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  58 T: 0.247E-01 Suc:   661 Eva:   705991 CPU:  0.138E+05 F: 6.1937 Pen:  0.04
30
 The best chi^2 values: 6.15068
 j:  59 T: 0.222E-01 Suc:   598 Eva:   725991 CPU:  0.142E+05 F: 6.1256 Pen:  0.01
73
 The best chi^2 values: 6.10831
 j:  60 T: 0.200E-01 Suc:   494 Eva:   745991 CPU:  0.146E+05 F: 6.0714 Pen:  0.01
53
 The best chi^2 values: 6.05616
 j:  61 T: 0.180E-01 Suc:   470 Eva:   765991 CPU:  0.150E+05 F: 6.0714 Pen:  0.01
53
 The best chi^2 values: 6.05616
 j:  62 T: 0.162E-01 Suc:   393 Eva:   785991 CPU:  0.154E+05 F: 6.0308 Pen:  0.02
64
 The best chi^2 values: 6.00438
 j:  63 T: 0.146E-01 Suc:   324 Eva:   805991 CPU:  0.158E+05 F: 6.0308 Pen:  0.02
64
 The best chi^2 values: 6.00438
 j:  64 T: 0.131E-01 Suc:   344 Eva:   825991 CPU:  0.162E+05 F: 6.0252 Pen:  0.00
27
 The best chi^2 values: 6.02251
 j:  65 T: 0.118E-01 Suc:   263 Eva:   845991 CPU:  0.165E+05 F: 6.0252 Pen:  0.00
27
 The best chi^2 values: 6.02251
```

ATV^{ORF273} - New Fold for a Thermo- and Acido-Stable Protein from the *Acidianus* Two-Tailed Virus

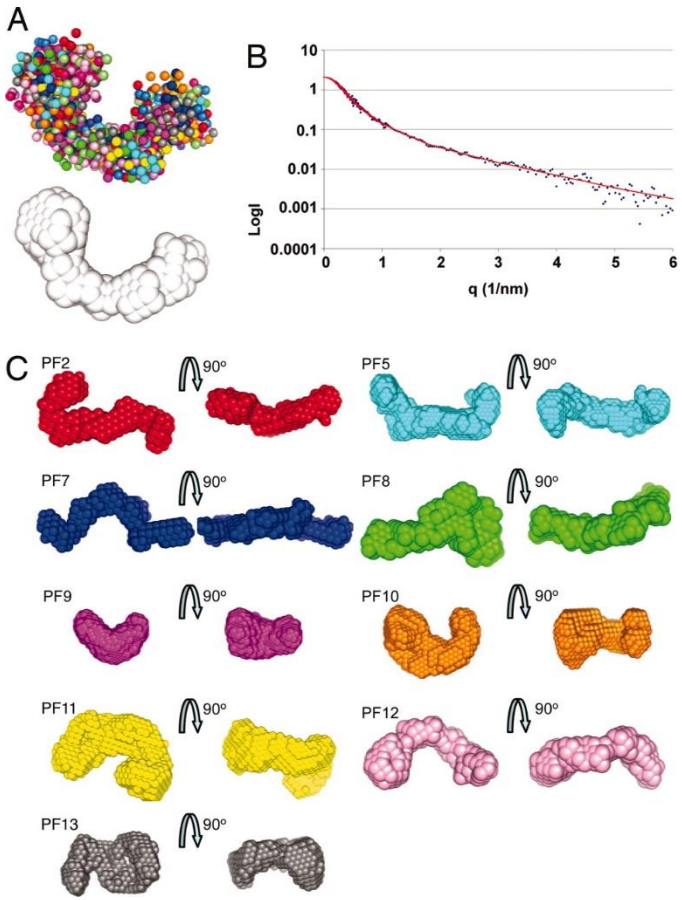


Three orthogonal views of each of the five *ab initio* envelopes calculated imposing (from left to right) no, binary, orthorhombic, trigonal or tetragonal symmetry.

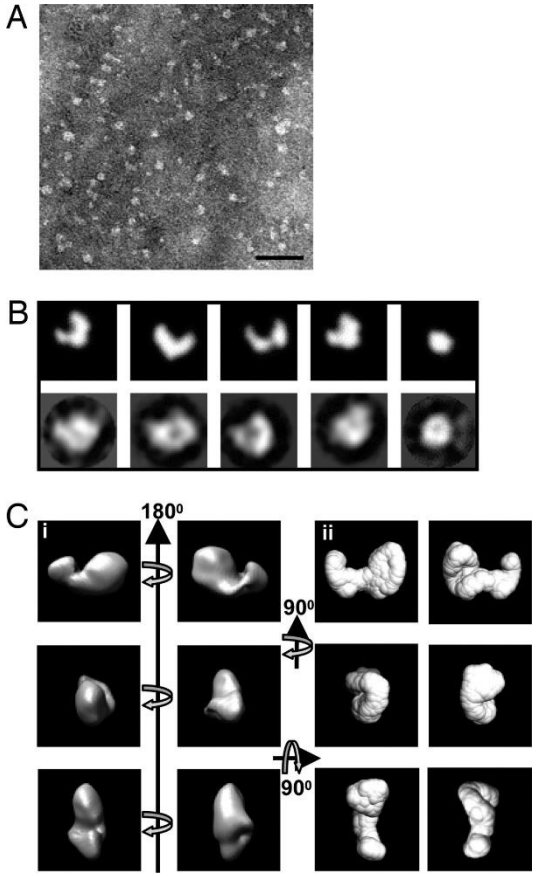
Fitting of the SASREF model obtained from the open dimer using P2 symmetry into the P222 DAMMIF model. The fitting was performed by the program Chimera

Felisberto-Rodrigues et al., PLoS One, 2012;7(10):e45847.

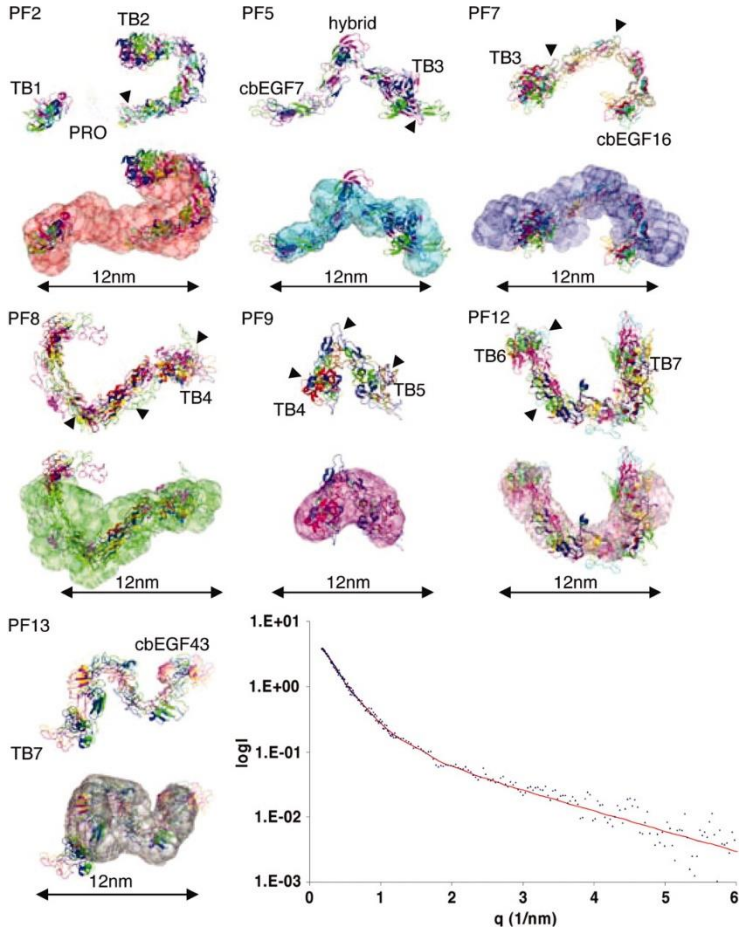
SAXS based Modeling to Determine Structure of Fibrillin-1



Structures generated *ab initio* from the SAXS data. Shapes were simulated *ab initio* by using the programs DAMMIN and GASBOR.



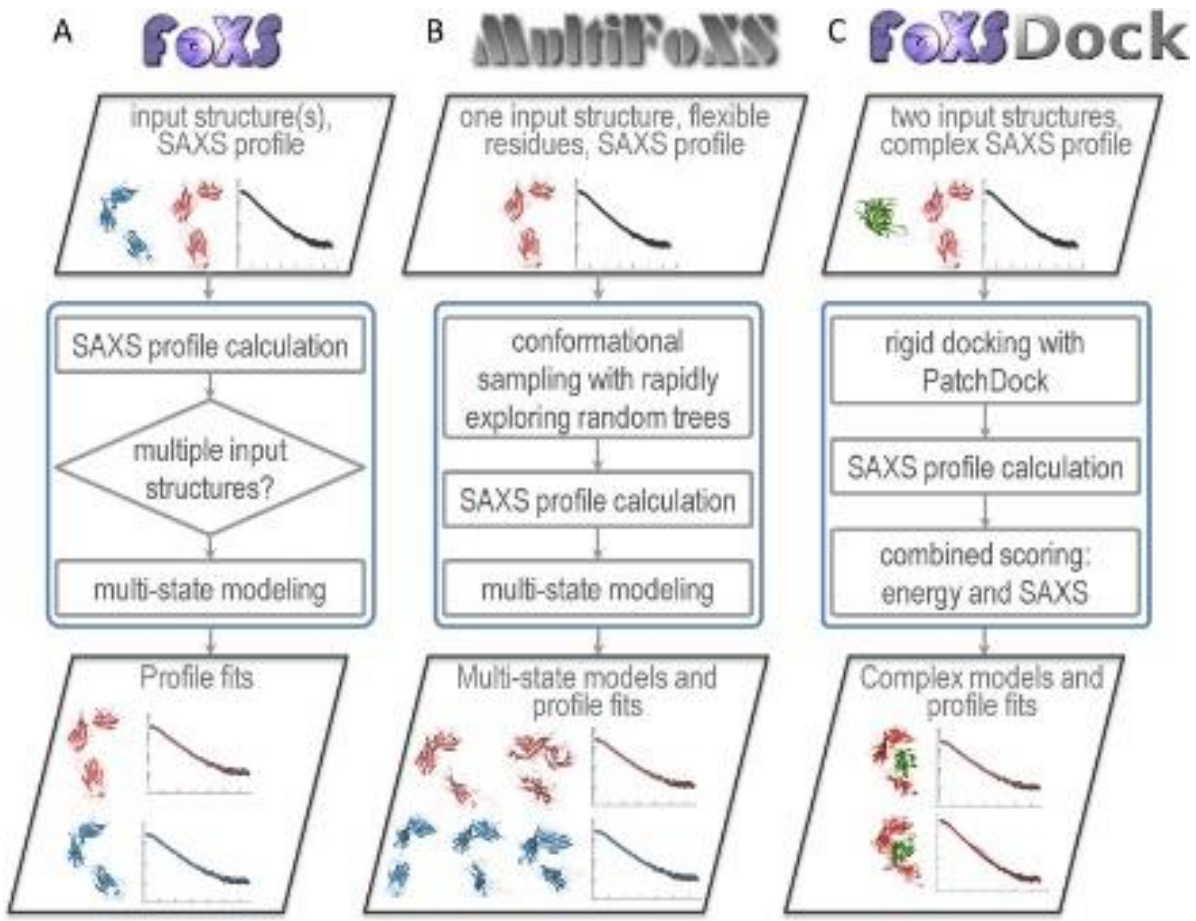
Comparison of the structures of PF11 determined by single-particle TEM and solution SAXS.

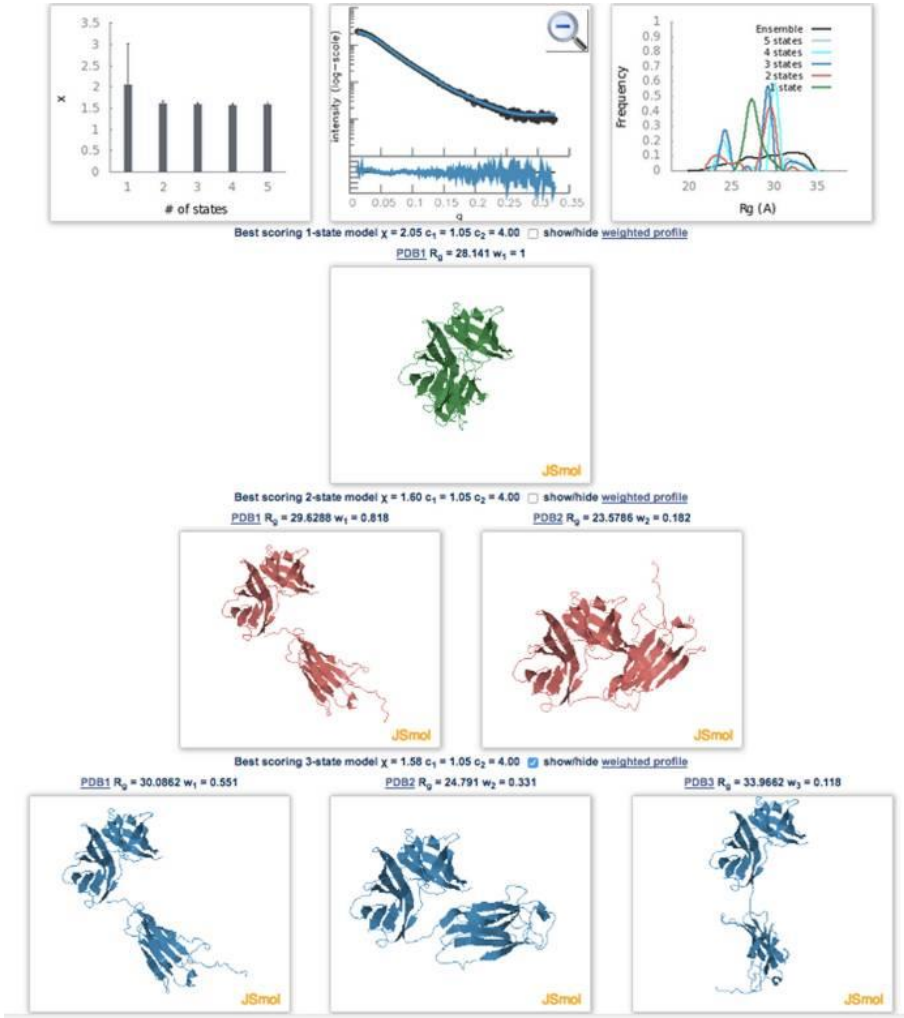


Rigid-body modeling was performed with the programs SASREF and BUNCH to the experimental SAXS data.

Clair Baldock et al. PNAS 2006;103:32:11922-11927

Fitting Multiple States against SAXS DATA





The unique capability of FoXS webserver is a possibility to account for multiple states contributing to a single observed SAXS profile.