SASREF: Global Rigid Body Modelling

- Starts from arbitrary initial positions and orientations of the subunits
- Employs simulated annealing to search interconnected arrangement of the subunits without clashes
- Random movement/rotation at one SA step
- Fitting the SAXS/SANS data by minimizing the target function

\[ E(X) = \sum \chi^2[I_{exp}(s), I(X,s)] + \sum \alpha_i P_i(X) \]

- Penalty terms describe additional restraints

Interconnectivity and steric clashes (natural restraints)

- Subunit arrangements with steric clashes and disconnected models are penalized.
- Overlap: $C_a-C_a$ distance < 4 Å.
Interconnectivity and steric clashes (natural restraints)

- To ensure the interconnectivity of the entire complex each subunit should have a contact with at least one other subunit.
- The contact distance between $C_\alpha$ atoms of distinct subunits: 4-7 Å.
Contacts restraints

- From binding affinity studies or from mutagenesis data the information on contacting subunits and even individual residues can be available.
- Such information is accounted for by specifying the ranges of residues or nucleotides which can be involved in interactions between the partners.
- Spring force potentials are added as penalties
Possible constraints

Symmetry

Groups Pn / Pn2 (n=1..6), P23, P432 and icosahedral symmetry can be taken into account.
- fewer spatial parameters to describe the model
- selection rules for the partial amplitudes:
  \( m \) equal to 0 or multiples of n,
  for Pn2, terms of order \( l0 \) with odd \( l \) and all imaginary parts vanish

Fixation of subset

Some subunits can be fixed at the initial positions and orientations to keep their mutual arrangement
Test examples

- Symmetric tetramer of pyruvate oxidase

- Dimeric & trimeric functional units of hemocyanin (Patrice)

- TK Met ectodomain - multidomain protein with no extended linkers
Modelling of multidomain proteins

MRGSHHHHHHH GSGVPSRVIH IRKLPIDVTE GEVISLGLPF GKVTNLLMLK
GKNQAFIEMN TEEAANTMVN YYTSVTPVLR GQPIYIQFSN HKELKTDSSP
NQARAQAALQ AVNSVQSGNL ALAASAAAAVD AGMAMAGQSP VLRIIVENLF
YPVTLDDLHQ IFSKFGTVLK IITFTKNNQF QALLQYADPV SAQHAKLSLD
GQNIYNAACCT LRIDFSKLTS LNVKYNNDKS RDYTRPDLPS GDSPQSLDQT
MAAAFGLSVP NVHGALAPLA IPSAAAAAAA AGRIAIPGLA GAGNSVLLVS
NLNPERVTPQ SLFILFGVYG DVQVRKILFN KKKNALVQMA DGNQAQLAMS
HLNGHKHLGK PIRITLSKHQ NVQLPREGQE DQGLTKDYGN SPLHRFKKPG
Modelling of multidomain proteins

- A combined approach is proposed to build the models of multidomain proteins with large and flexible interdomain linkers.
- The latter are represented as DR chains which are attached to the appropriate terminals in rigid domains.
- A single modification of a model is a rotation about one or two randomly selected DR(s).
Modelling of multidomain proteins
Building native-like folds of linkers

Absence of steric clashes

Neighbors distribution along the sequence

Loop compactness may also be required

$$Rg_{id} = 3^3 \sqrt{n_l}$$
Simultaneous fitting of multiple data sets from deletion mutants
PTB protein

MRGSHHHHHH GS GVPSRVIH IRKLPIDVTE GEVISLGLPF GKVTNLLM
GKNQAFIEMN TEEAANTMVN YYTSVTPVLR GQPIYIQFSN HKELKTDSSP
NQARAQAALQ AVNSVQSGNL ALAASAAAADV AGMA MAGQSP VLRRIIVENF
YPVTLDVLHQ IFSKFGTVLK IITFTKNQF QALLQYADPV SAGHAKEDO
GQNIYNACCT LRIDFSKLTS LNVKYNNDKS RDYTRPDLPS GDSQPSLST
MAAFA GLSVP NVHGALAPLA IPSAAAAAAA AGRIAIPG LA AGNSVLLVS
NLNPERVTPQ SLFILFGVYG DVQRVKILFN KKlenaLVQMA DGNQAQLAK
HLNGHK LHGK PIRTLSKHQ NVQLPREGQE DQGLTKD YGN SPLHRFKK
SKNFQ NIPPP SATLHLSNIP PSVSEEDLKV LFSSNGG VVKA GFKFFQKDKR
MALIQ MGSVE EAVQALIDLH NHDLGENHHL RVSFSKSTI
CORAL: new in ATSAS 2.4
Sasref vs Bunch
Merging Sasref and Bunch

CORAL = COmplexes with RAndom Loops
RanLog library
Simplified input (config file)

NTER 20
A1.pdb
LINK 25
A2.pdb
CTER 10

NTER 45
B1.pdb

C1.pdb
Example

• GST-DHFR fusion protein (Kate)