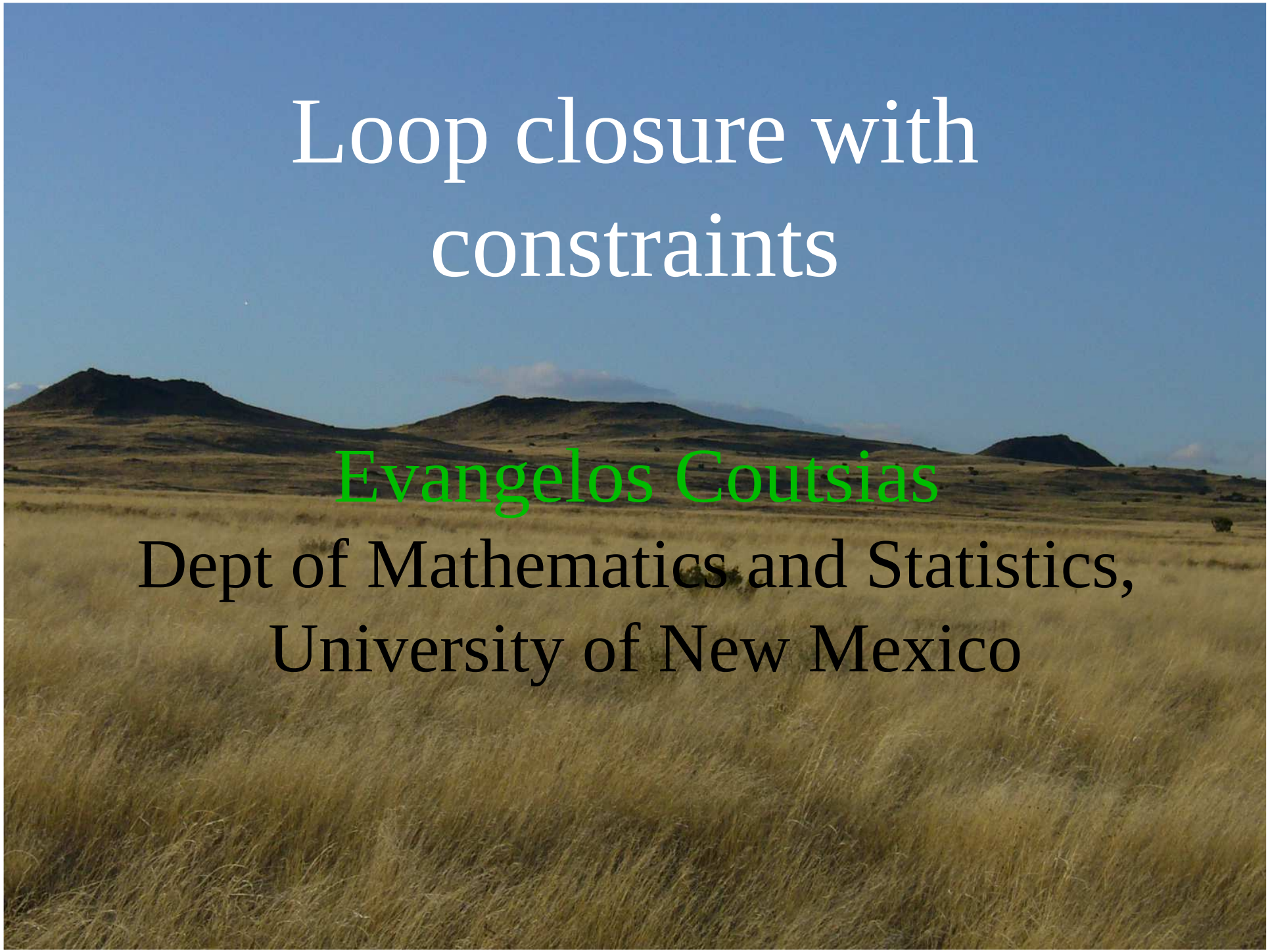
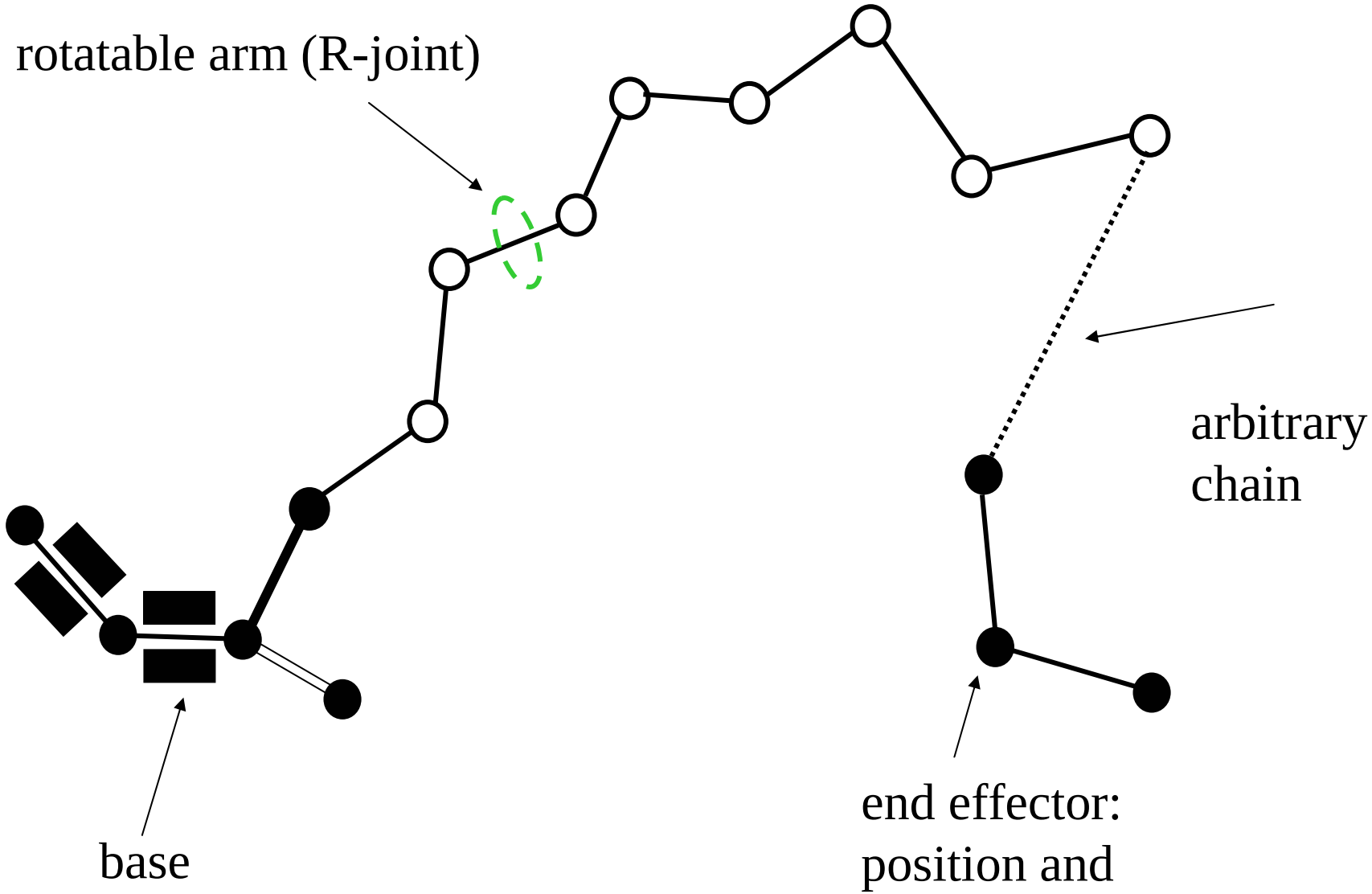


Loop closure with constraints

Evangelos Coutsias
Dept of Mathematics and Statistics,
University of New Mexico

outline

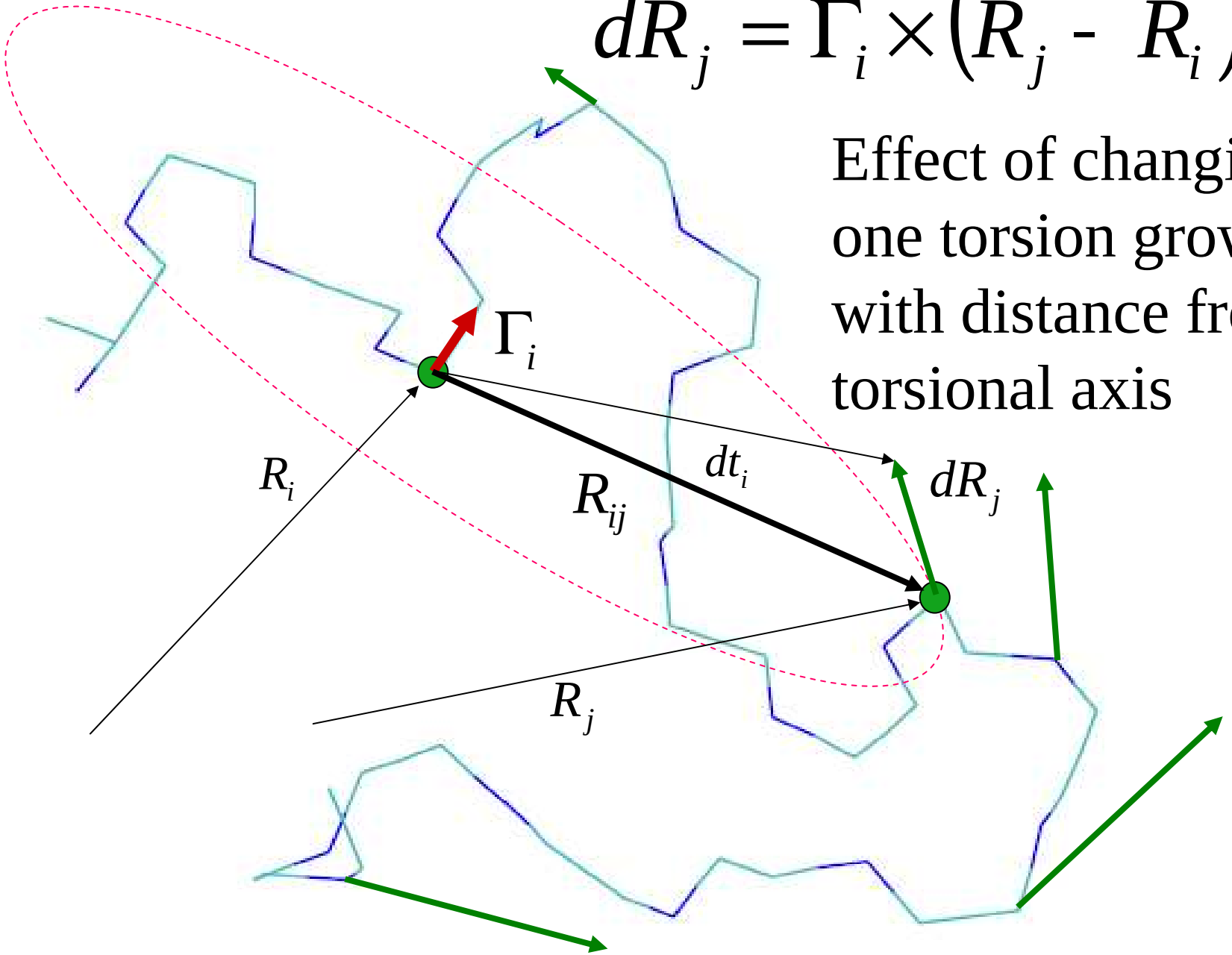
- Inverse Kinematics
- Tetrahedral equ & Bricard Octahedra
- Triaxial loop closure
- Conformational searches: complications
- Non-generic flexibility = failure
- Constraints: deterministic, approximate
- Example: fixed position/sidechain orientation



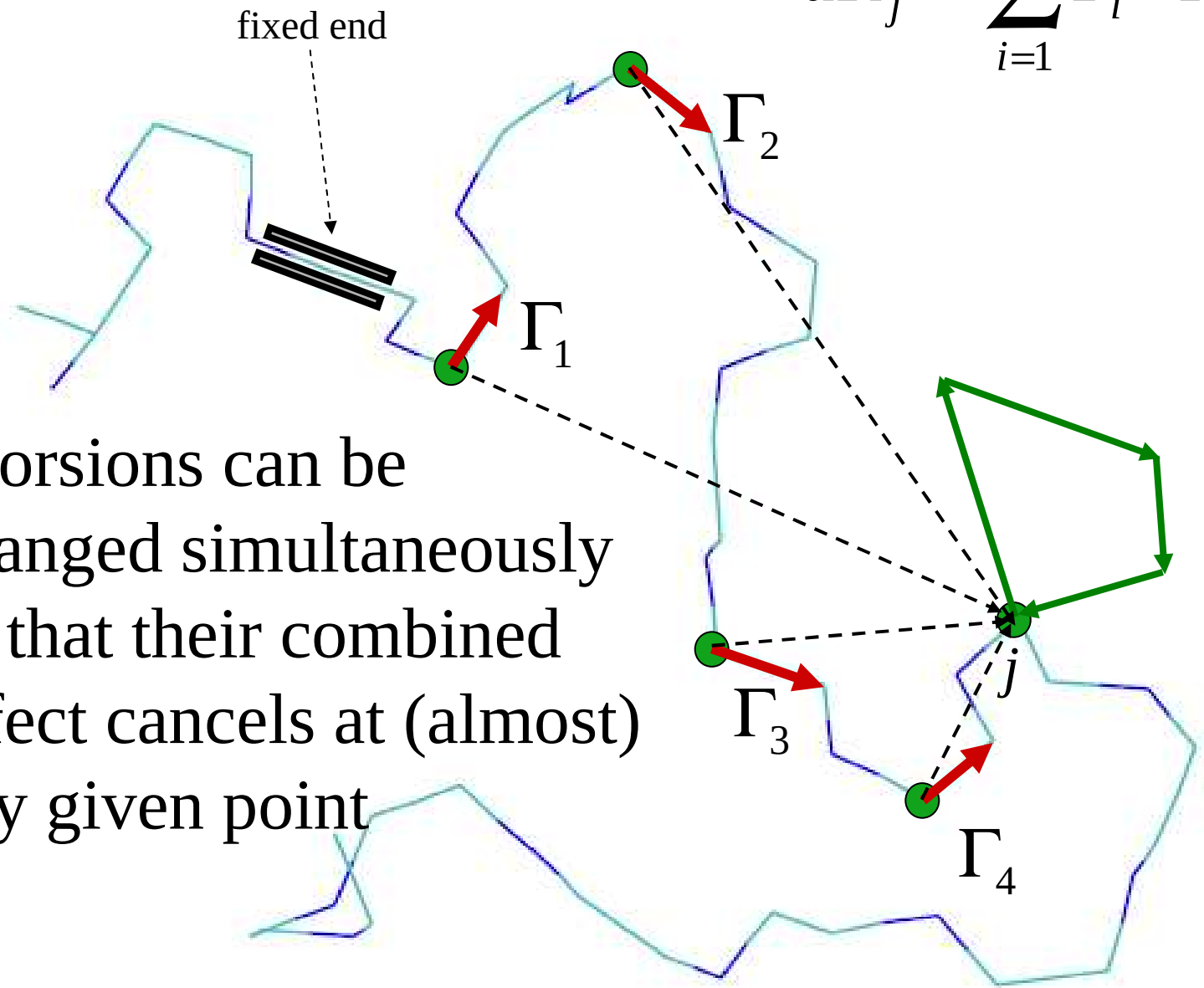
Inverse kinematics: given base, find torsions that place end effector

$$dR_j = \Gamma_i \times \left(R_j - R_i \right) dt_i$$

Effect of changing one torsion grows with distance from torsional axis

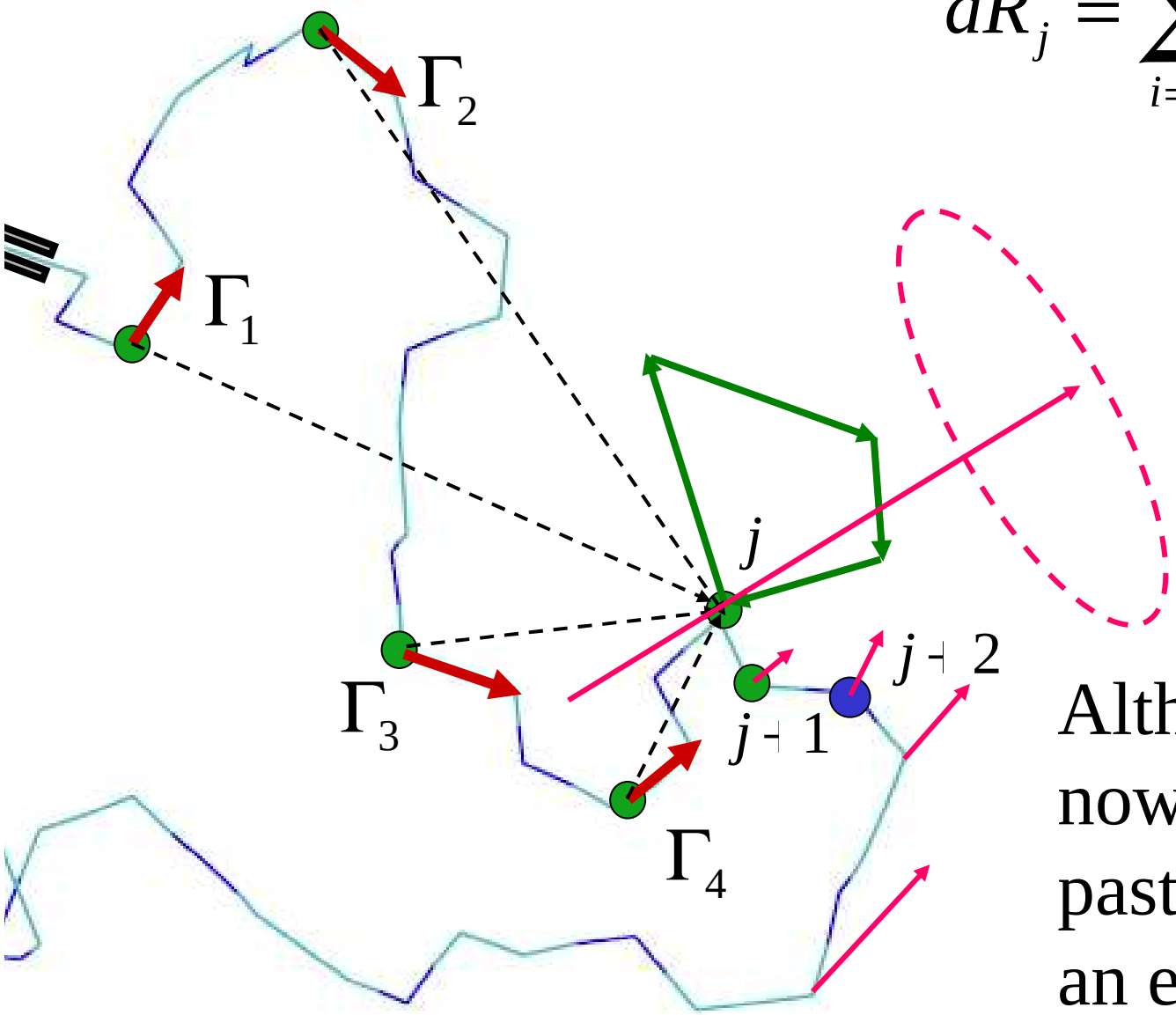


$$dR_j = \sum_{i=1}^{4} \Gamma_i \times R_{ij} dt_i = 0$$

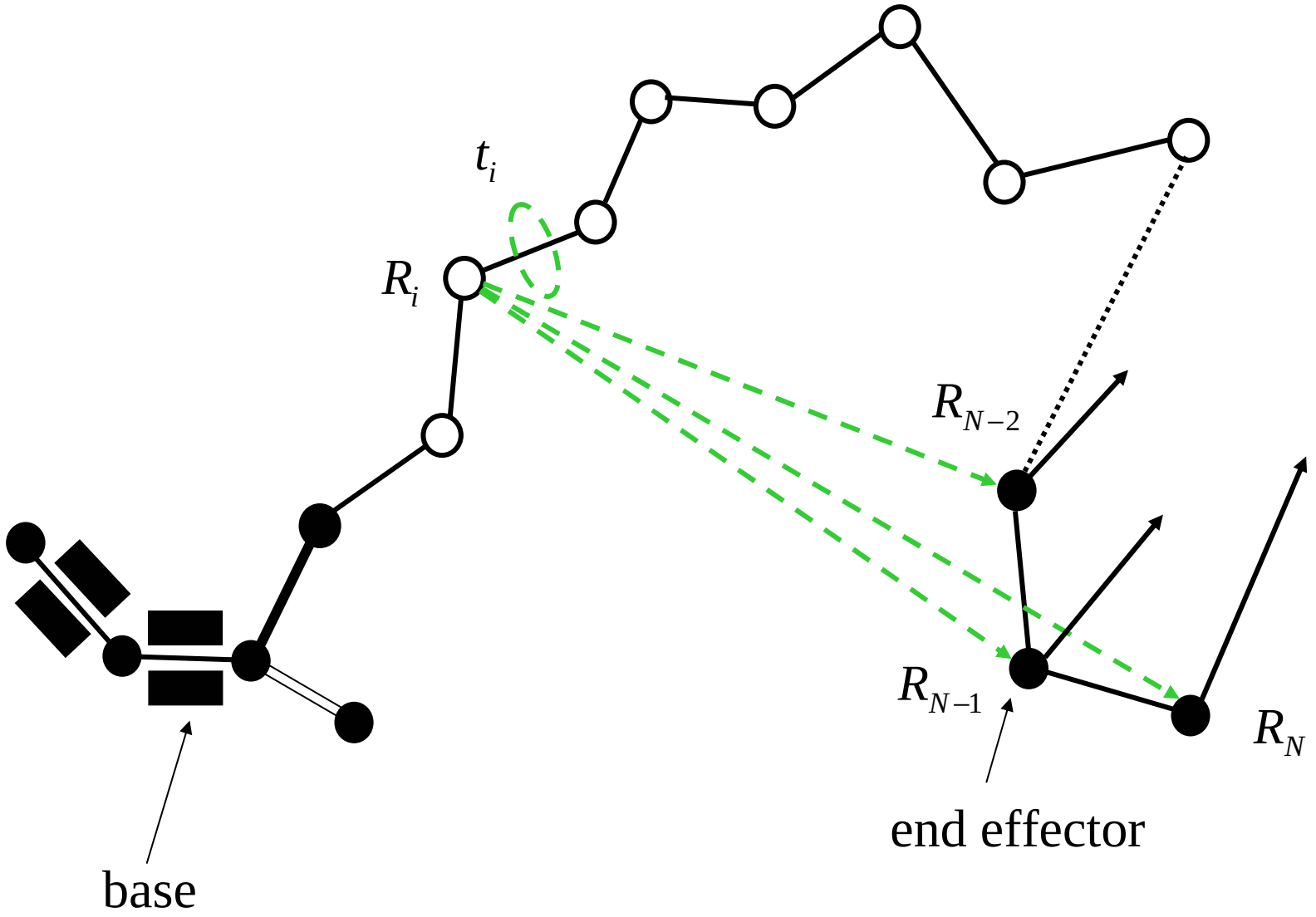


4 torsions can be changed simultaneously so that their combined effect cancels at (almost) any given point

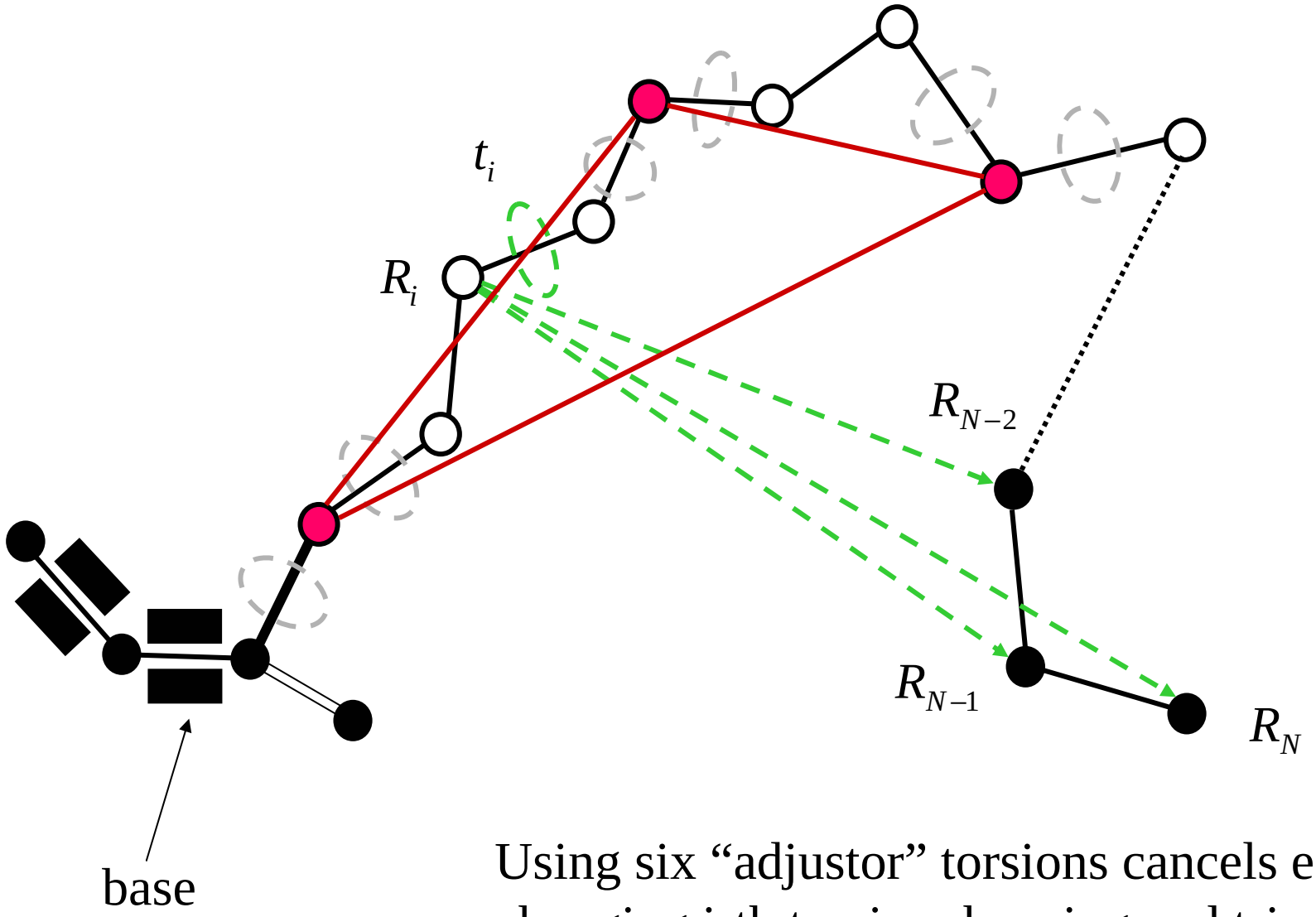
$$dR_j = \sum_{i=1}^{4} \Gamma_i \times R_{ij} dt_i = 0$$



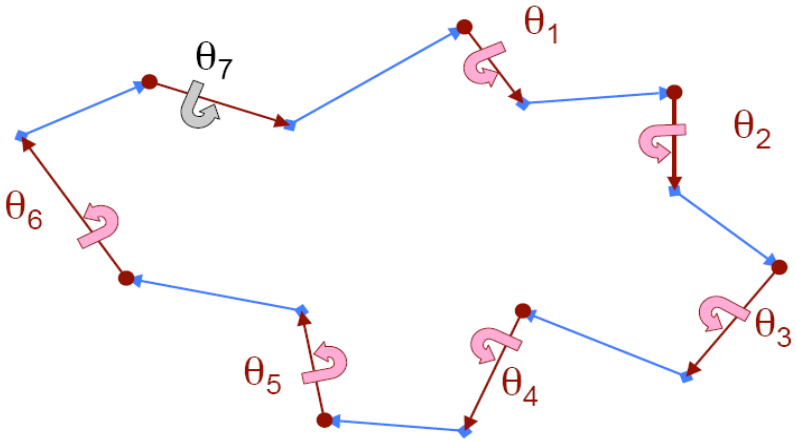
Although atom j is now fixed, the chain past j rotates about an effective axis through j



The end triad is moved due to change in the i-th torsion



Using six “adjustor” torsions cancels effect of changing i-th torsion, keeping end triad (and all subsequent atoms) fixed in space

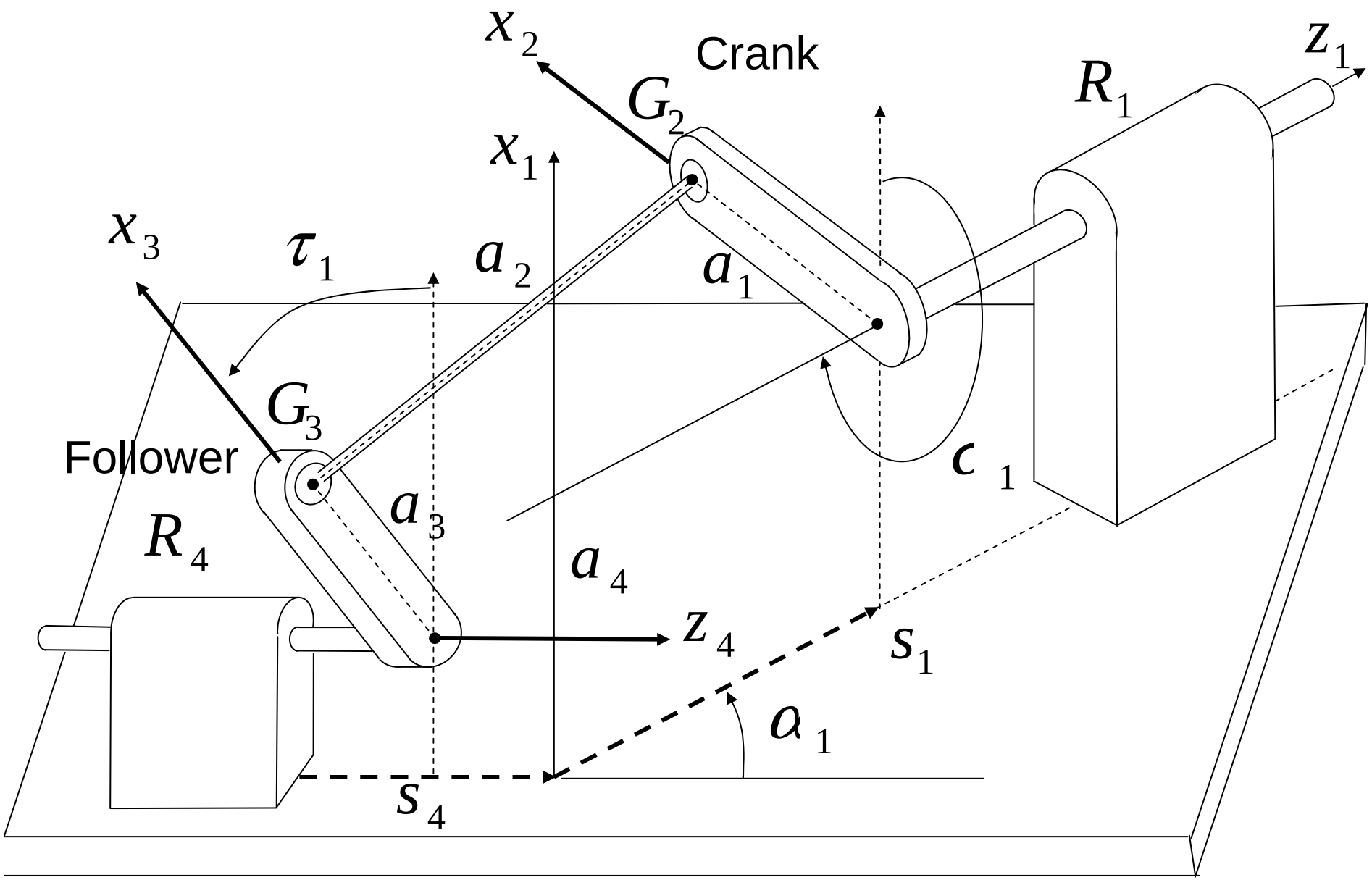


"The Mt. Everest of kinematics"

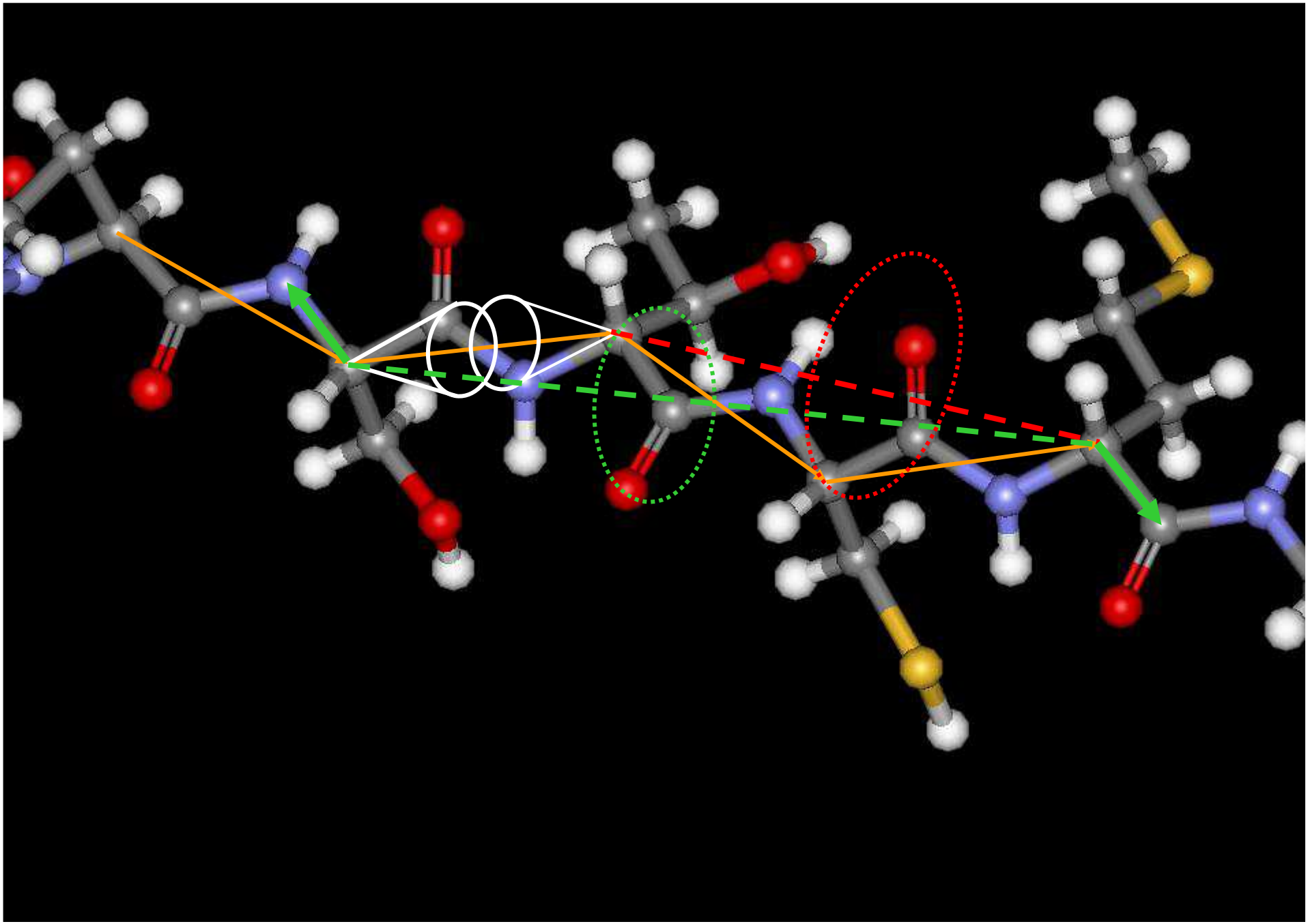
A. Freudenstein

Robotics 7R/7bar problem: Lee&Liang, 1988

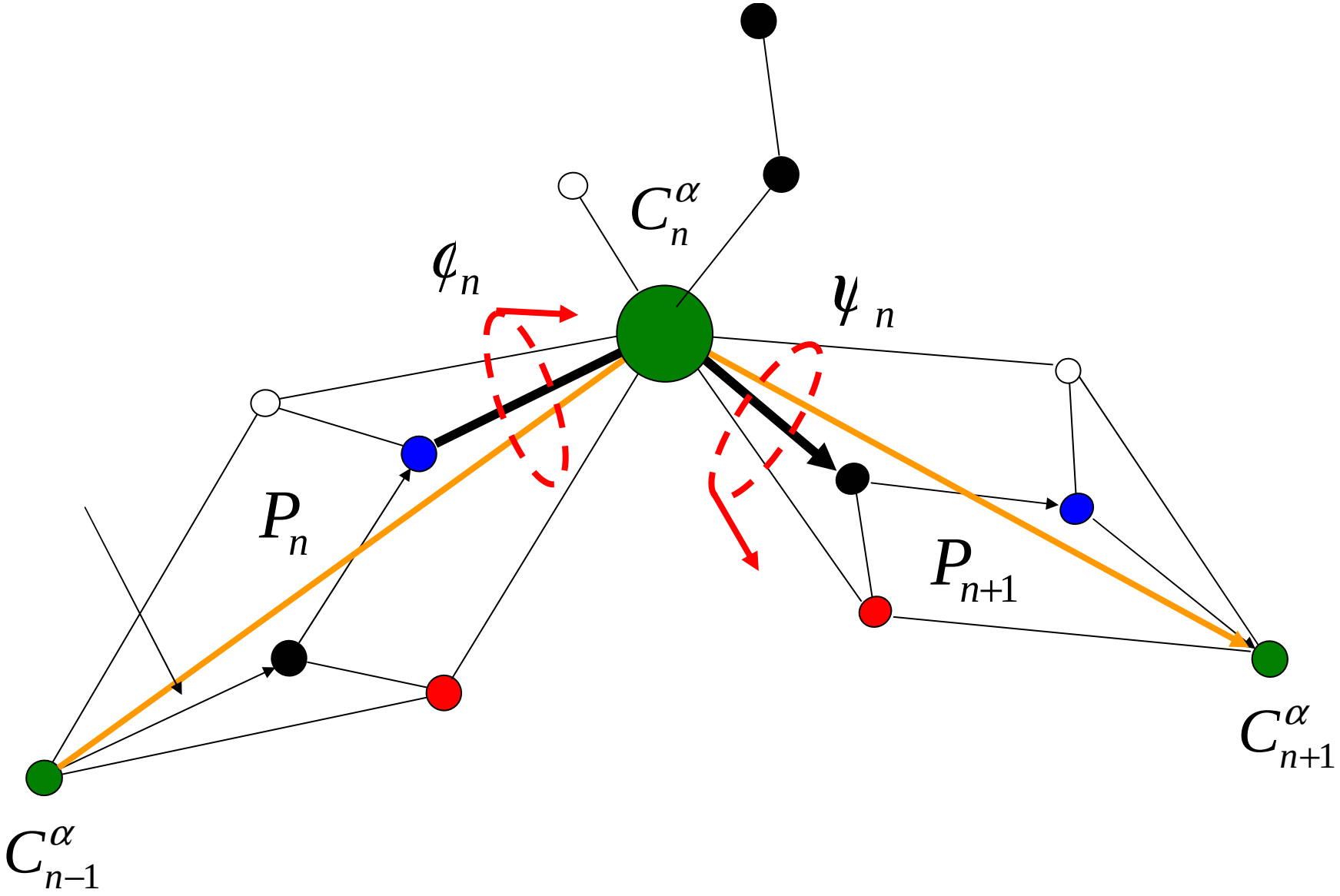
The formulation is quite involved; however it eventually leads to a generalized eigenproblem. The numerical computation in this form is approx. 10-100x slower than for the Triaxial-Dixon algorithm.



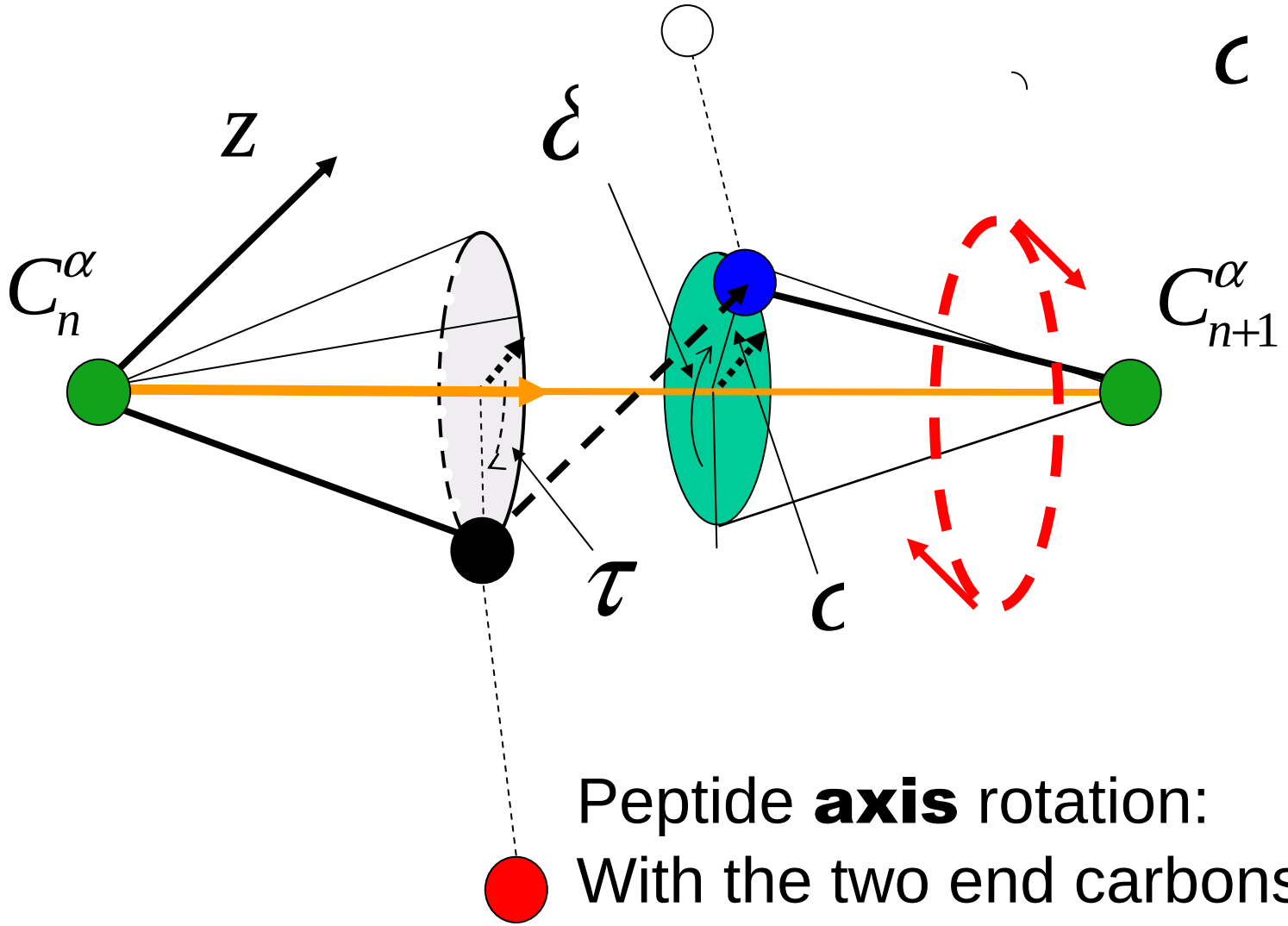
Two-revolute, two-spherical-pair mechanism



Study of localized motions in a polypeptide chain



The pep-2 “capstone”

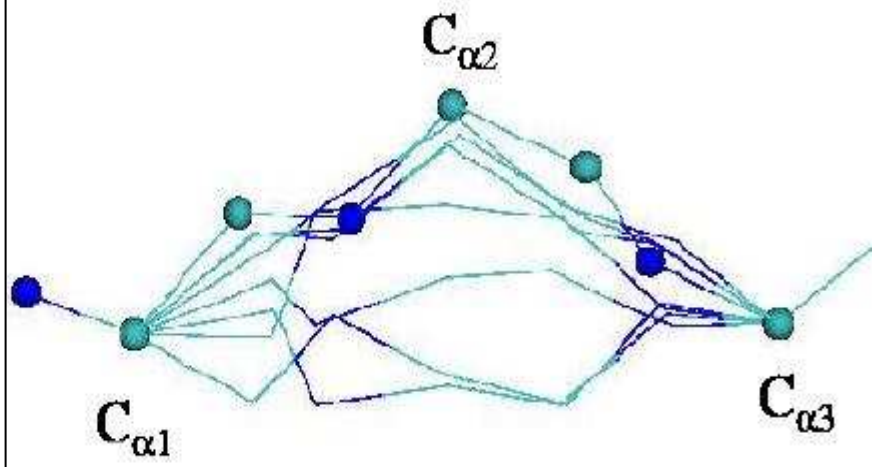


Peptide **axis** rotation:

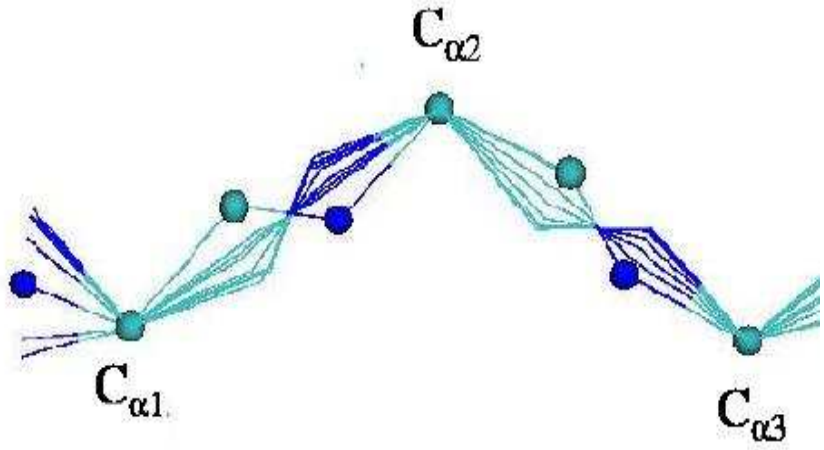
With the two end carbons fixed in space, the peptide unit rotates about the virtual bond

Representation of Loop Structures

In the space frame



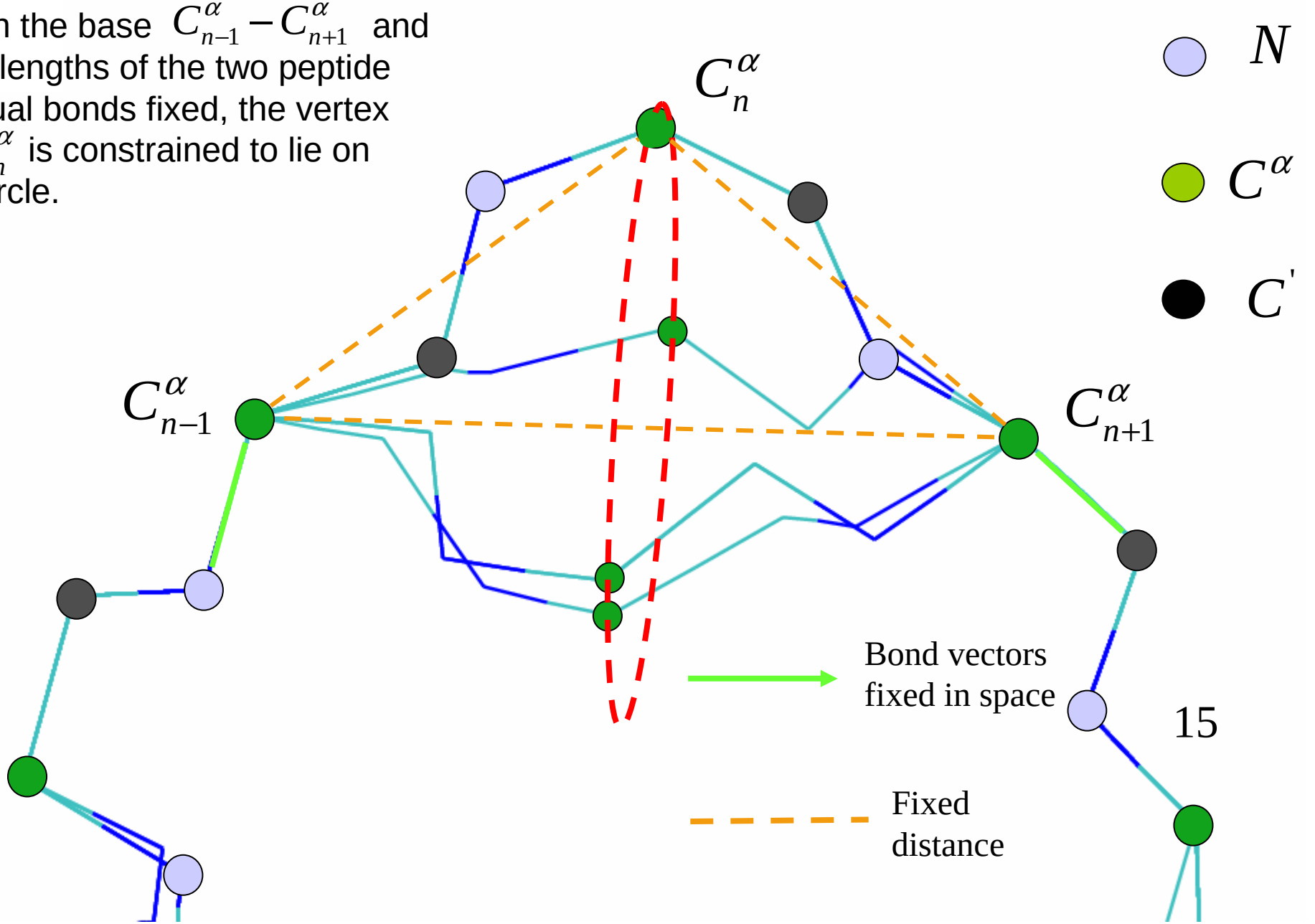
In the body frame

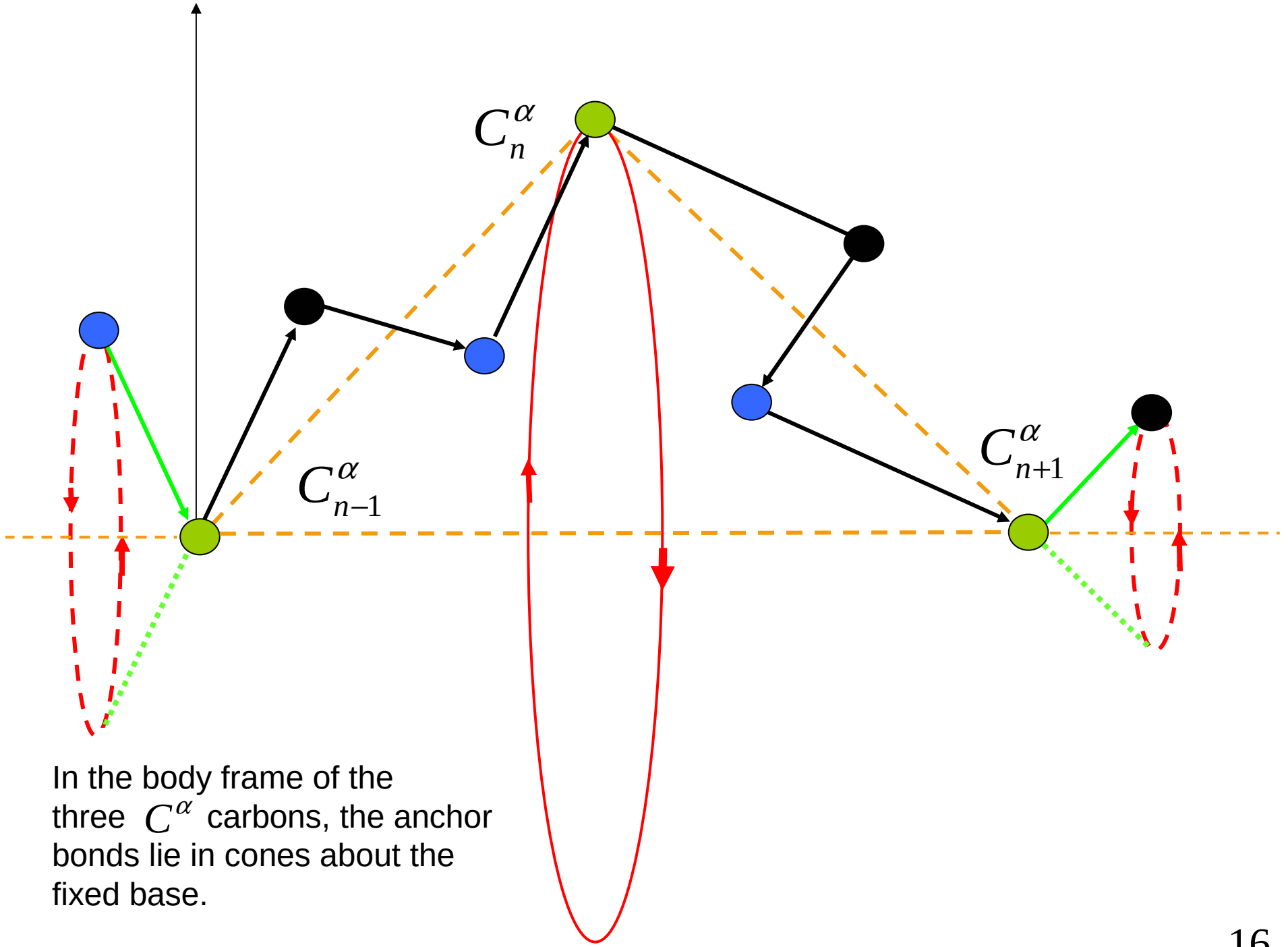


Coutsias, Seok, Jacobson, Dill, JCC 2004

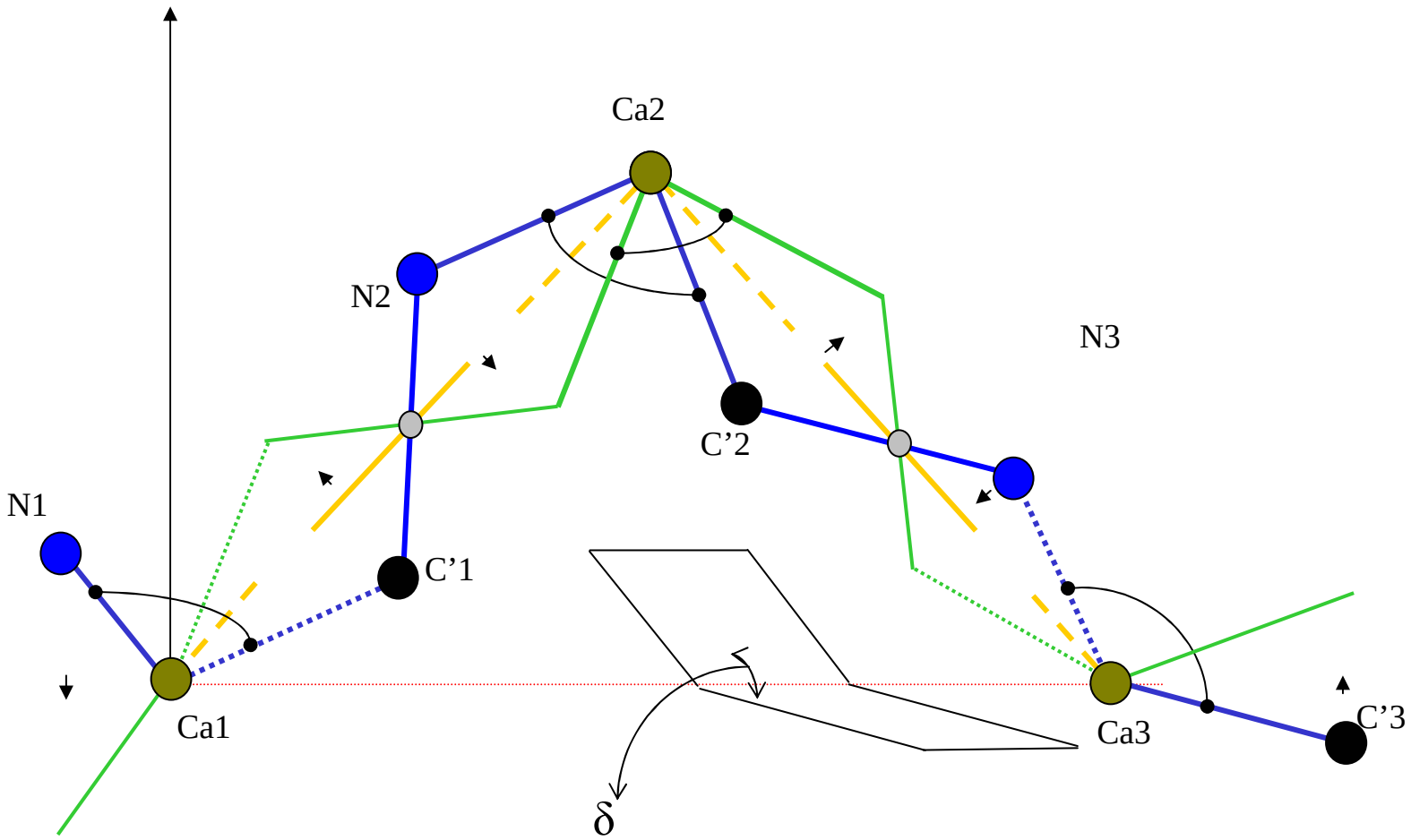
Tripeptide Loop Closure

With the base $C^{\alpha}_{n-1} - C^{\alpha}_{n+1}$ and the lengths of the two peptide virtual bonds fixed, the vertex C^{α}_{n} is constrained to lie on a circle.

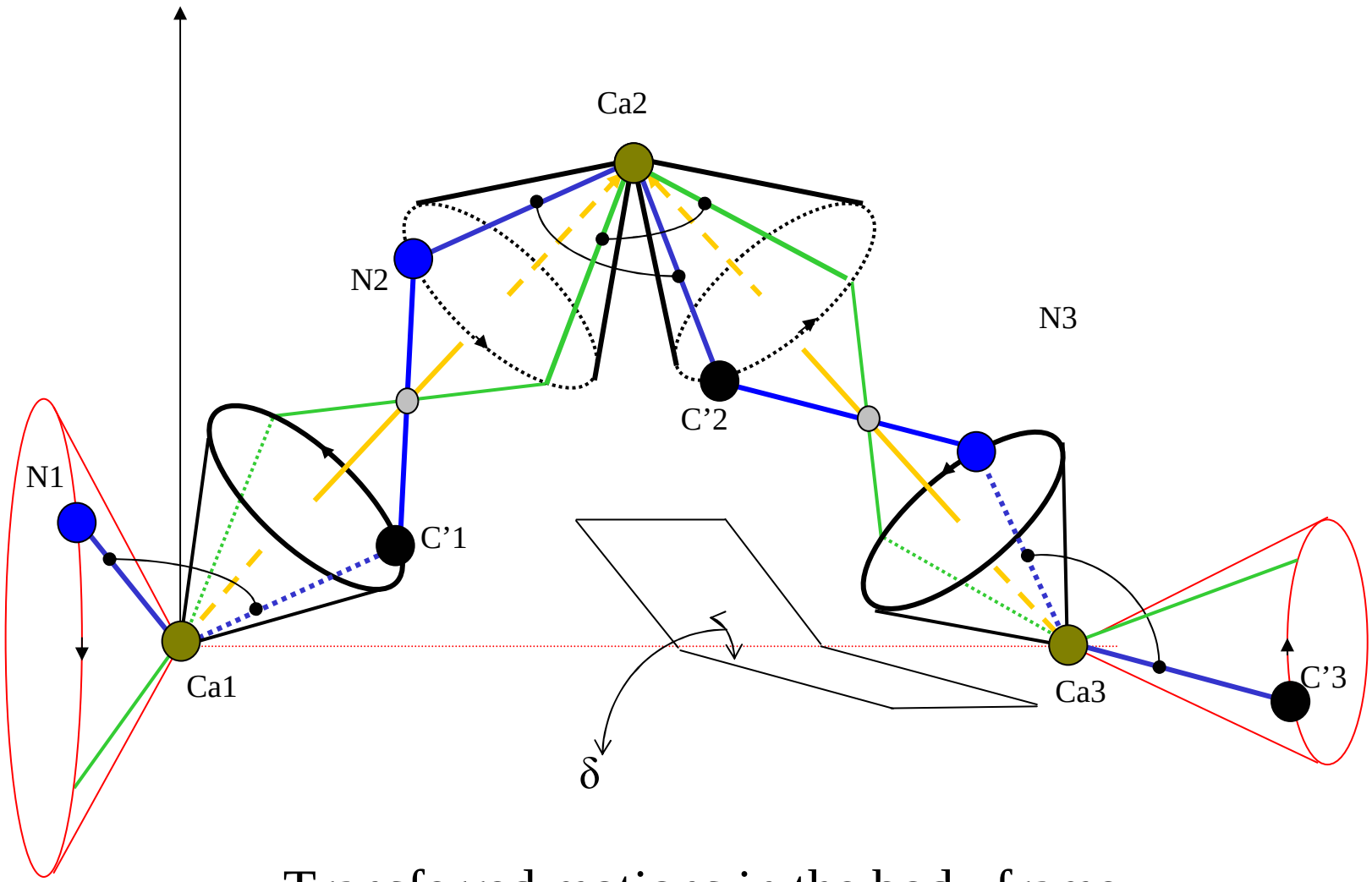




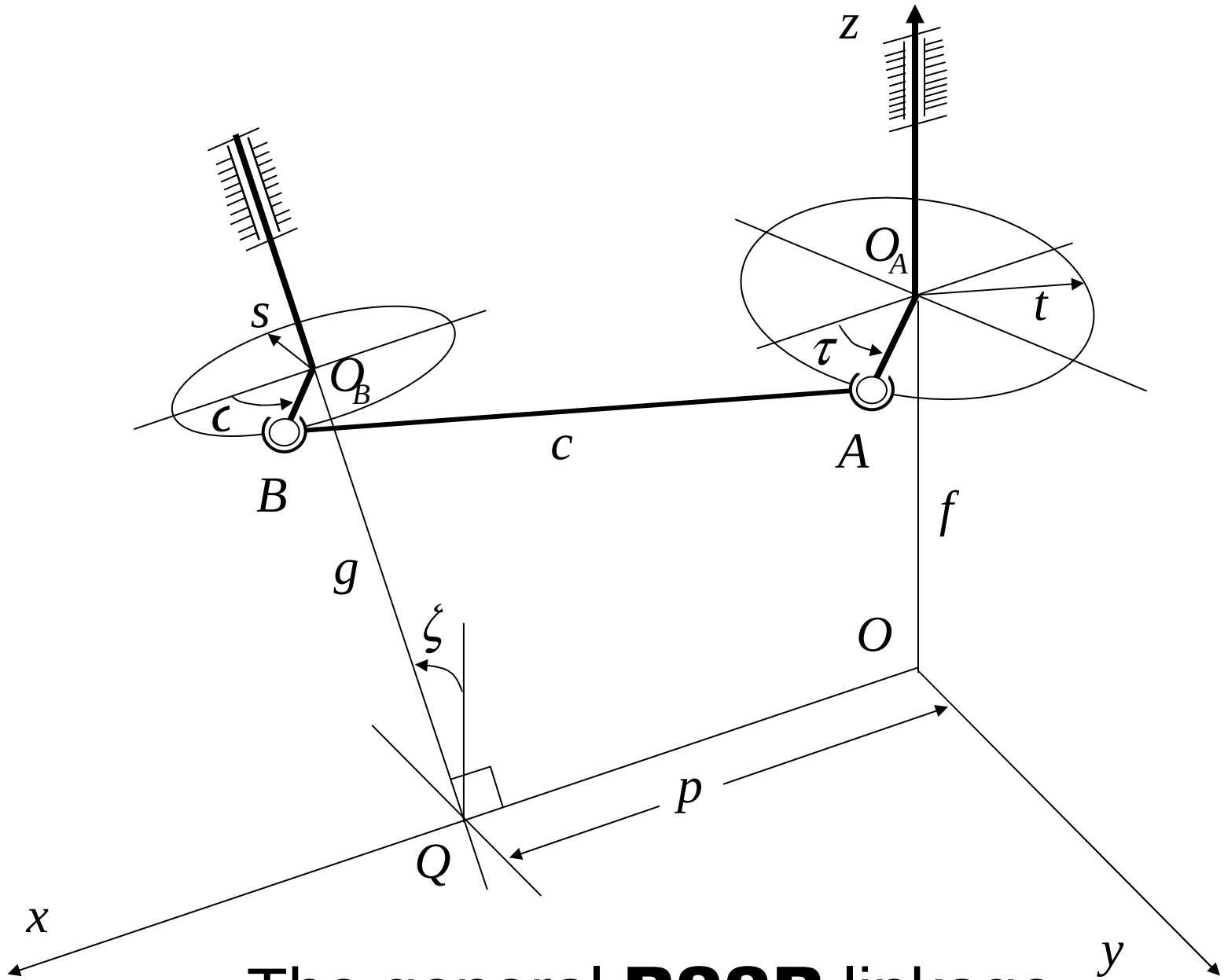
In the body frame of the three C^α carbons, the anchor bonds lie in cones about the fixed base.



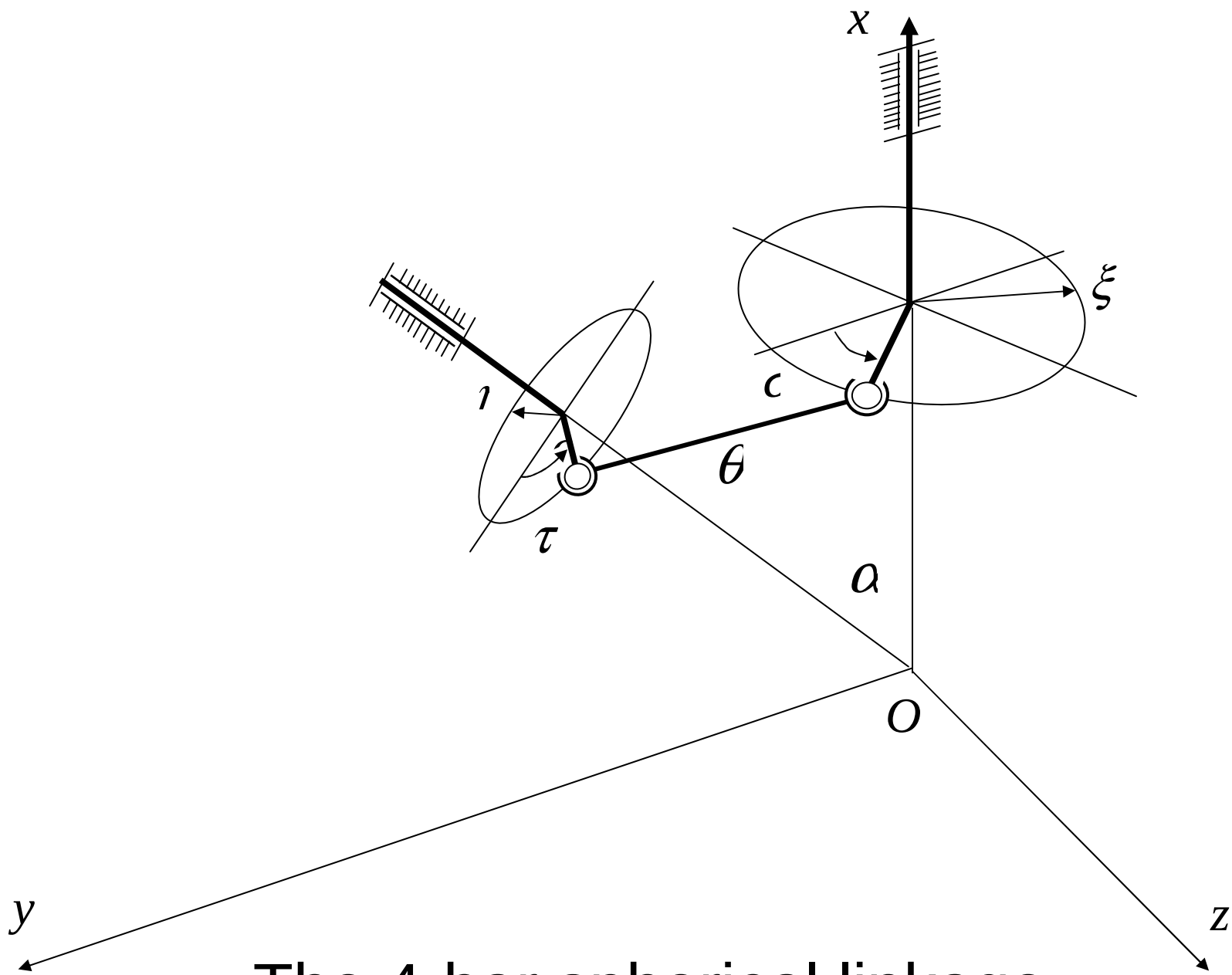
LOOP CLOSURE: find all configurations with two end-bonds fixed
The angle between the planes N1-Ca1-Ca3 and Ca1-Ca3-C’3 is given,
the orientation of the two fixed bonds (N1-Ca1 and Ca3-C’3) wrt the plane Ca1-Ca2-Ca3 can assume several values (at most 8 solutions are possible)



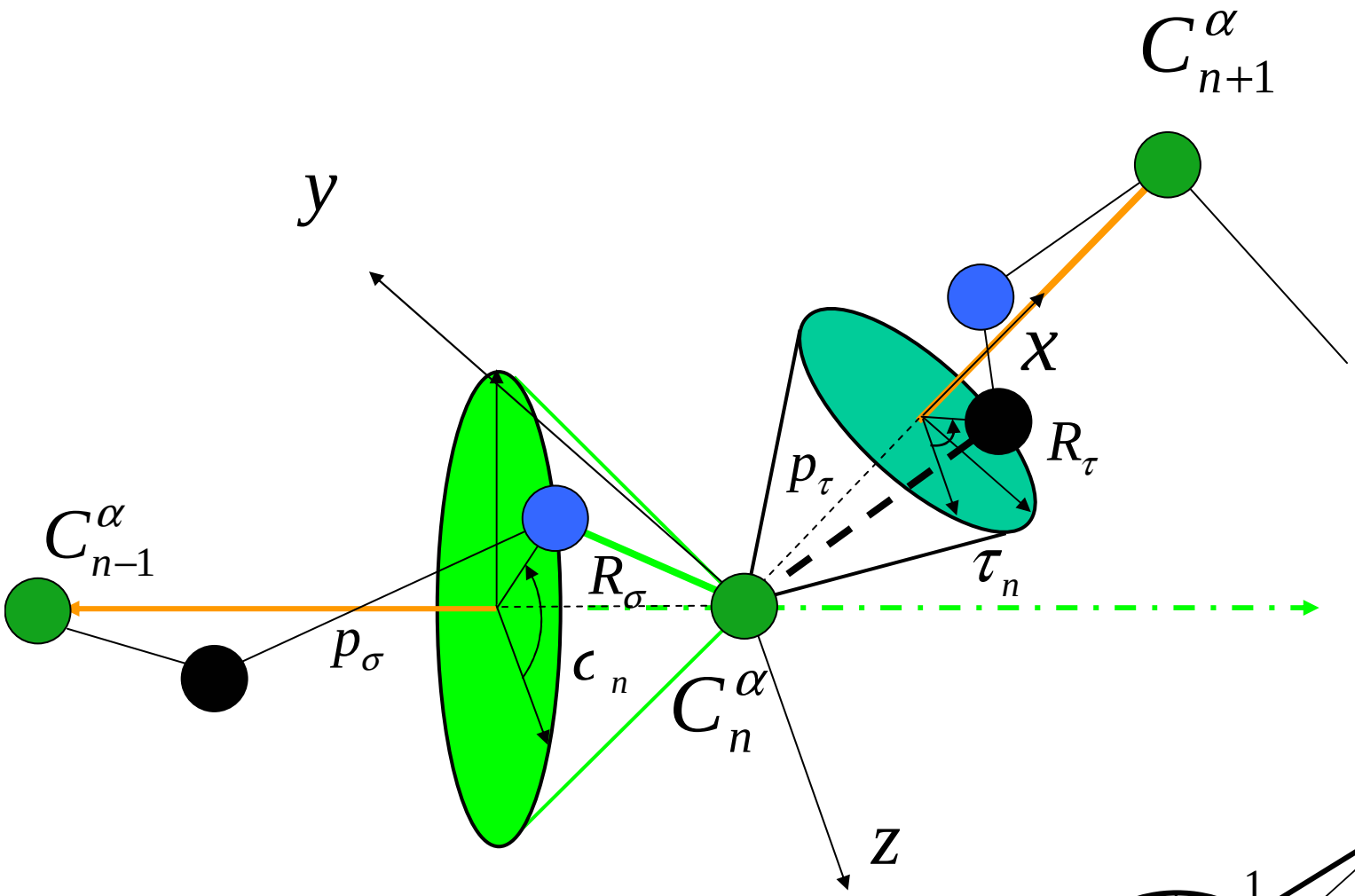
Transferred motions in the body frame
of three contiguous Ca carbon units:
In this frame the Ca carbons resemble
spherical 4-bar linkage joints

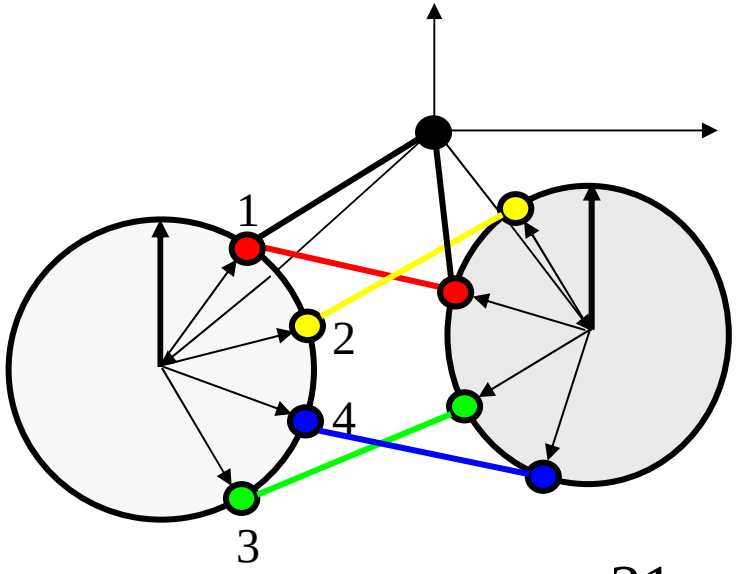


The general **RSSR** linkage

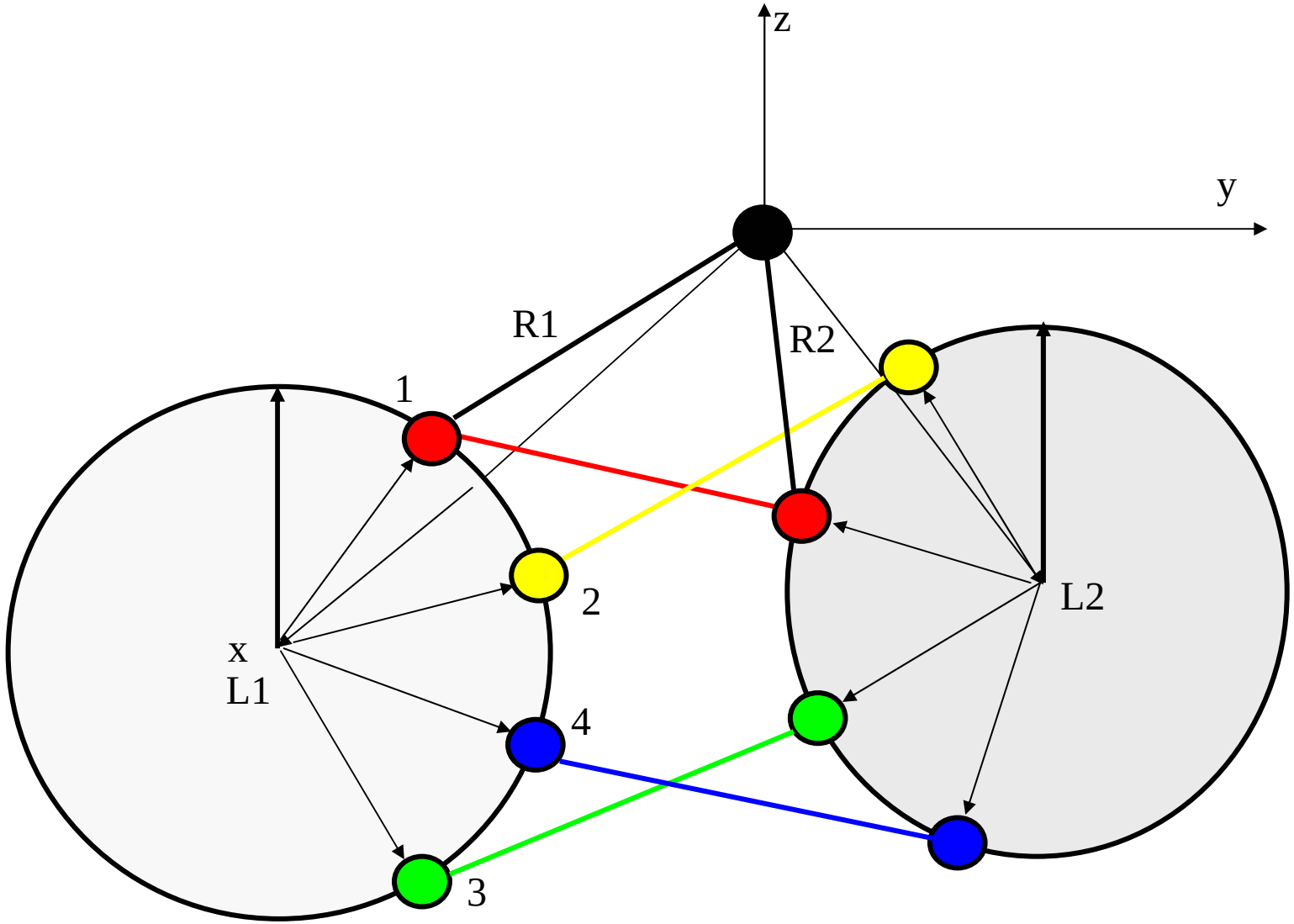


The 4-bar spherical linkage





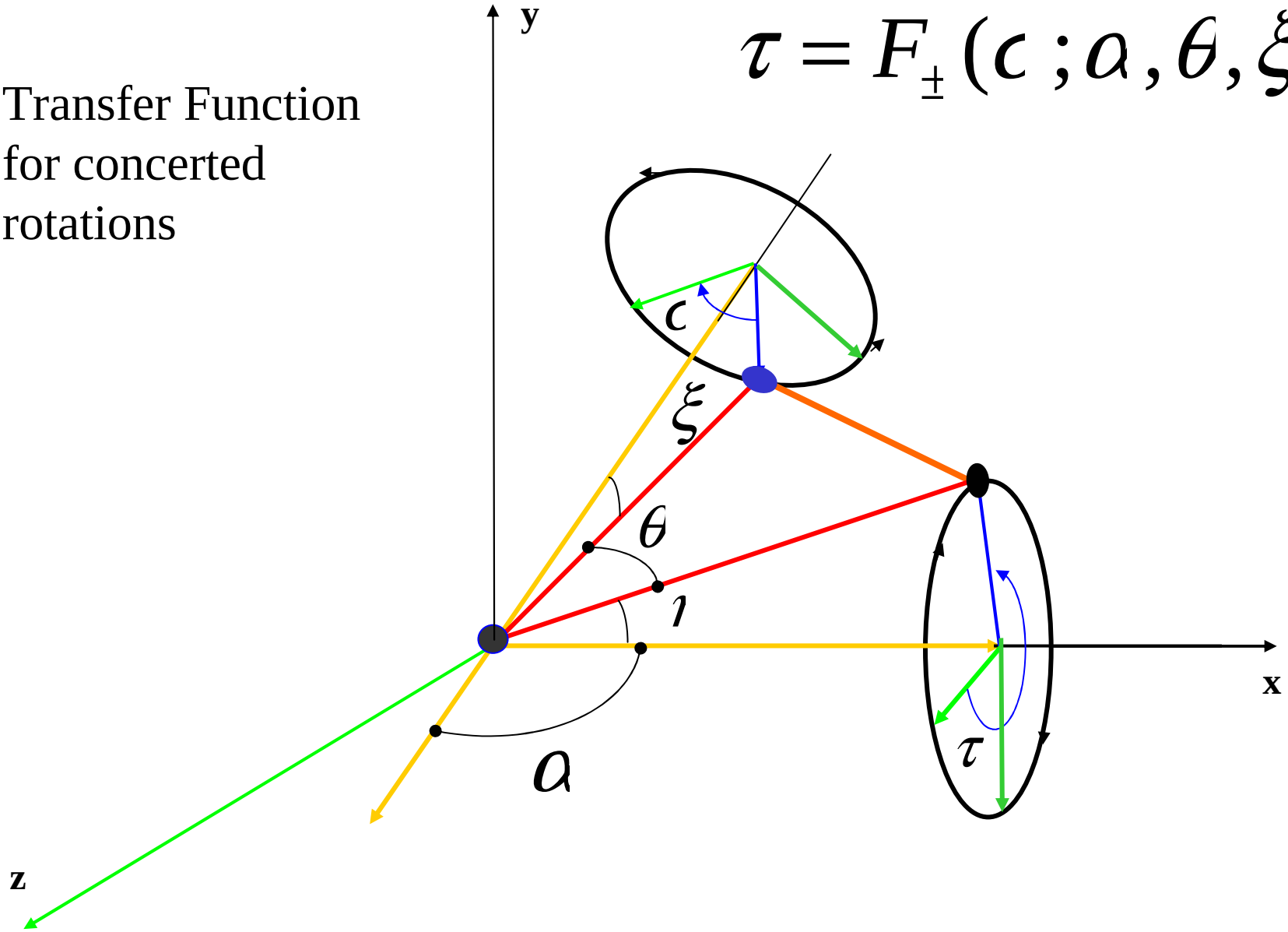
Coordinated rotation at junction of 2 rotatable bonds (the angle between the two bonds remains fixed as each rotates about its own peptide unit virtual axis).



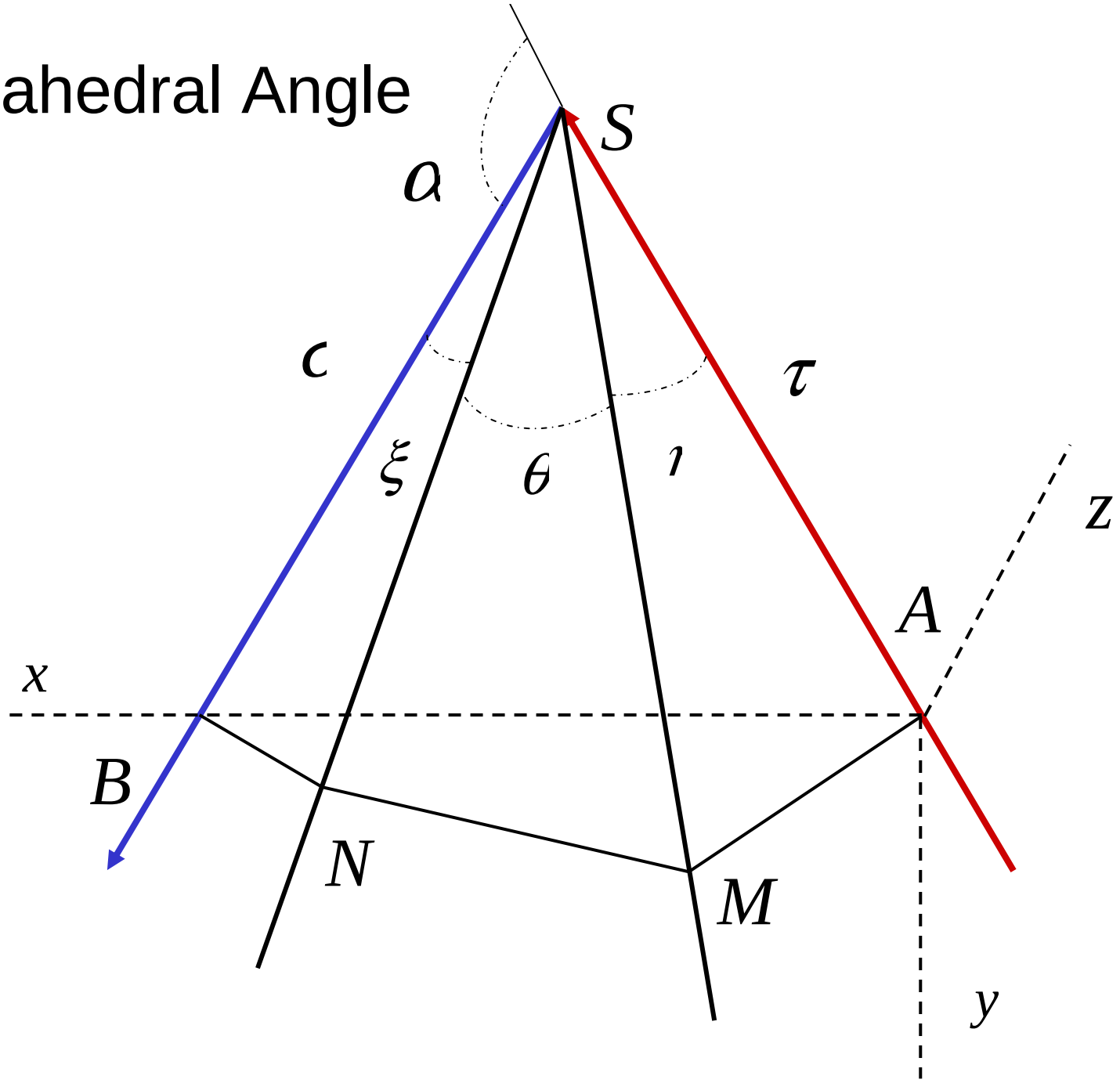
A complete cycle through the allowed values for φ (dihedral (R1,R2) -(L1,R1))and ψ (dihedral (R1,R2)-(L2,R2))

Transfer Function for concerted rotations

$$\tau = F_{\pm}(c\,;a,\theta,\xi,\imath\,)$$



Tetrahedral Angle



With angles a, x, h, q fixed, the torsions s, t are coupled.

Analysis is carried out by “rationalizing” trigonometric expressions through the half-angle formulas:

$$w = \tan\left(\frac{\sigma}{2}\right) \Leftrightarrow \cos\sigma = \frac{1-w^2}{1+w^2}; \sin\sigma = \frac{2w}{1+w^2}$$

$$u = \tan\left(\frac{\tau}{2}\right) \Leftrightarrow \cos\tau = \frac{1-u^2}{1+u^2}; \sin\tau = \frac{2u}{1+u^2}$$

Equation of the Tetrahedral Angle

Raoul Bricard, 1897

(study of flexible octahedra)

$$\left(Au^2+B\right)w^2+Cuw+\left(Du^2+E\right)=0$$

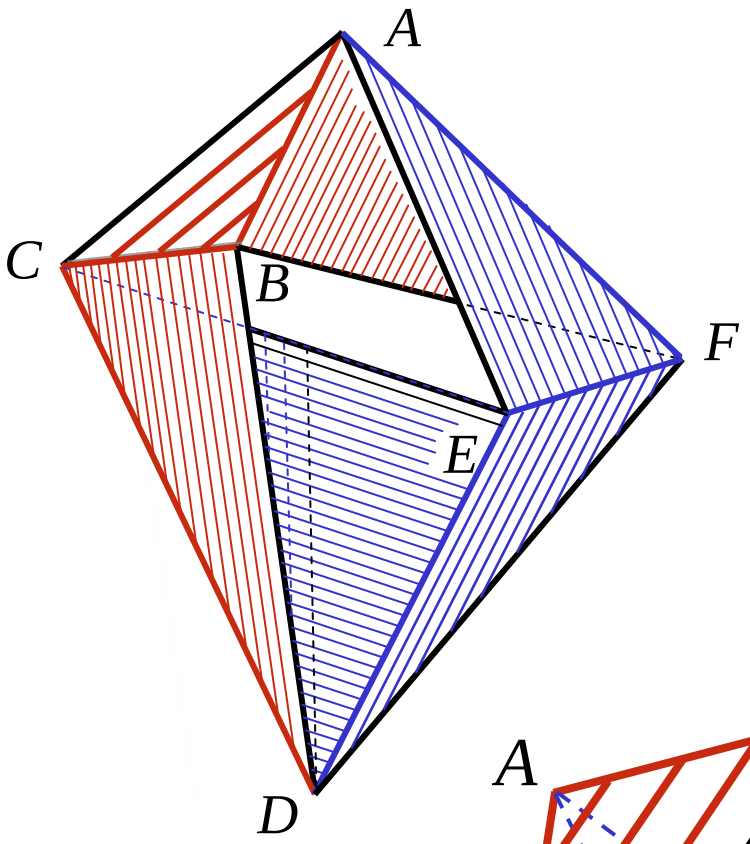
$$A=\cos\theta+\cos(\alpha-\xi-\eta)$$

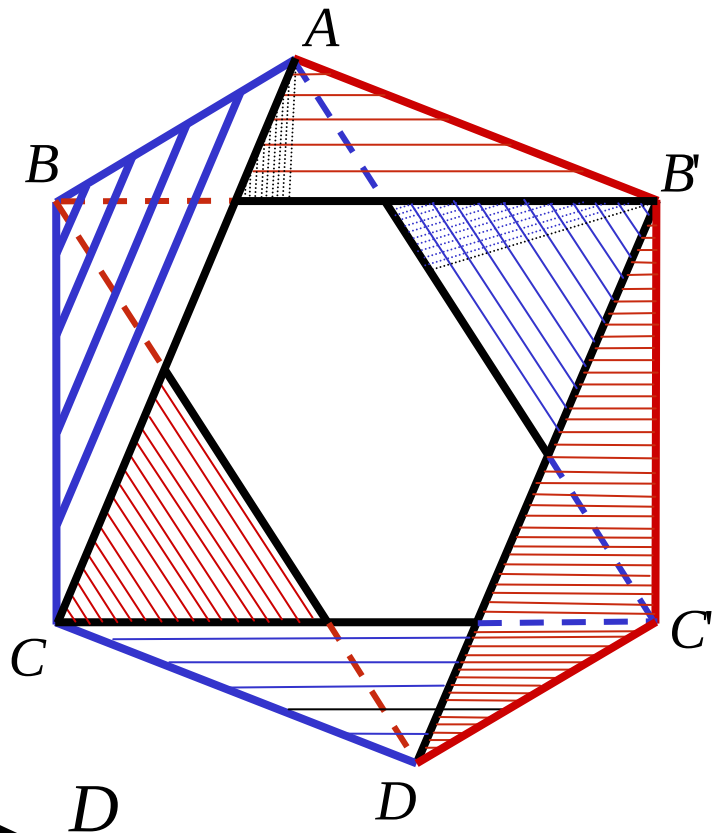
$$B=\cos\theta+\cos(\alpha+\xi-\eta)$$

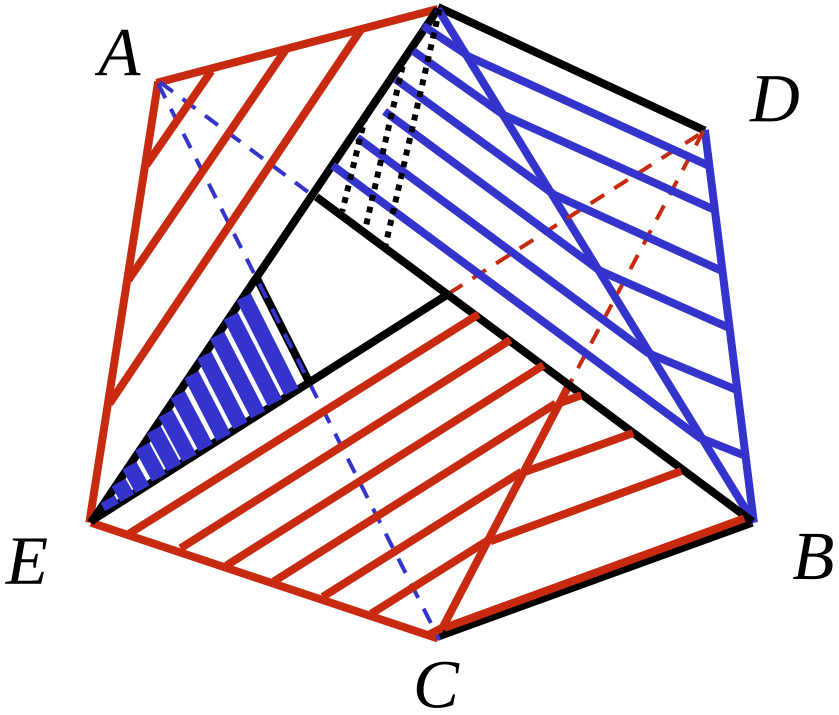
$$C=-4\sin\xi\sin\eta$$

$$D=\cos\theta+\cos(\alpha-\xi+\eta)$$

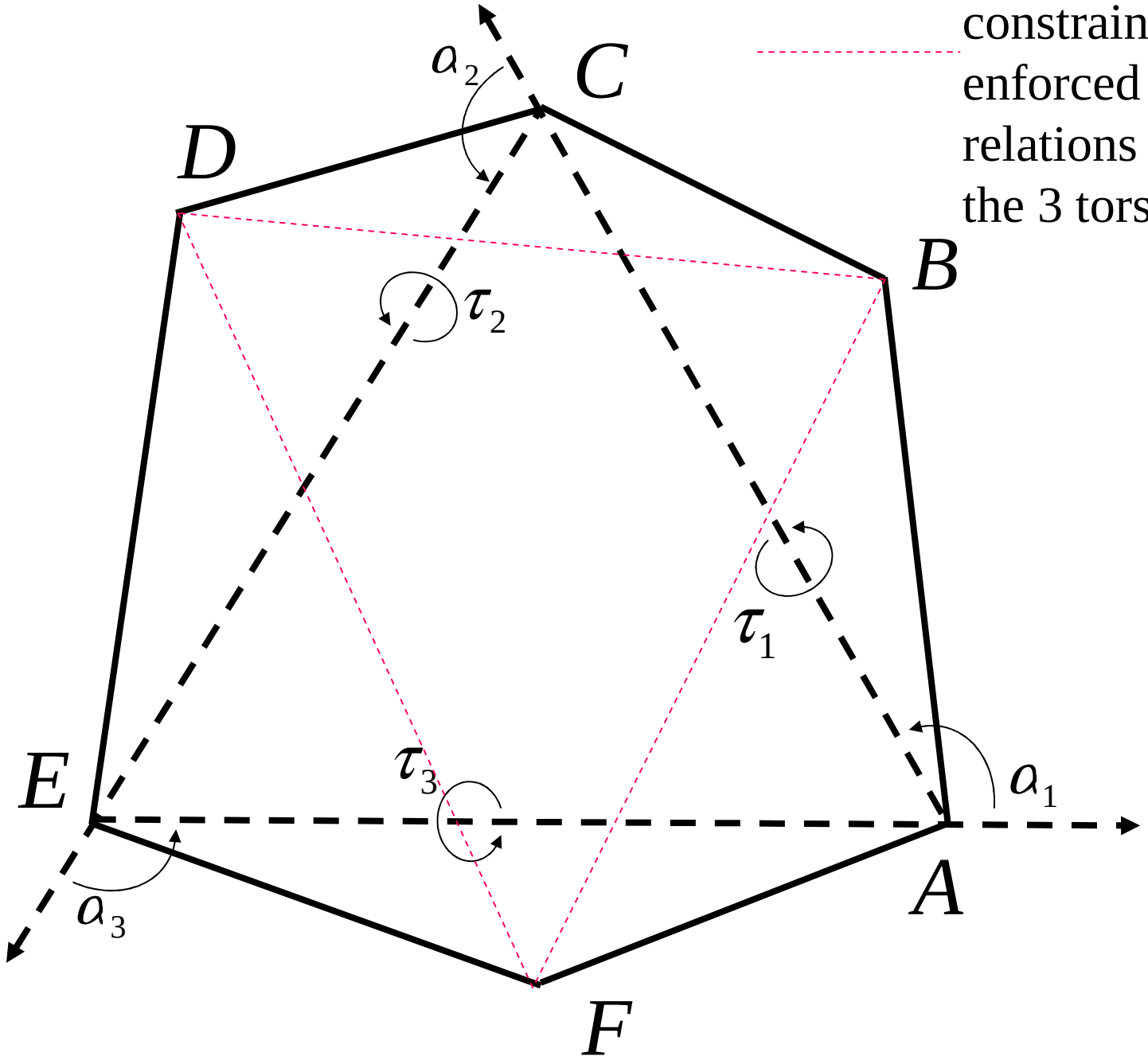
$$E=\cos\theta+\cos(\alpha+\xi+\eta)$$







Bricard
Flexible
Octahedra



constraints: enforced as relations among the 3 torsions

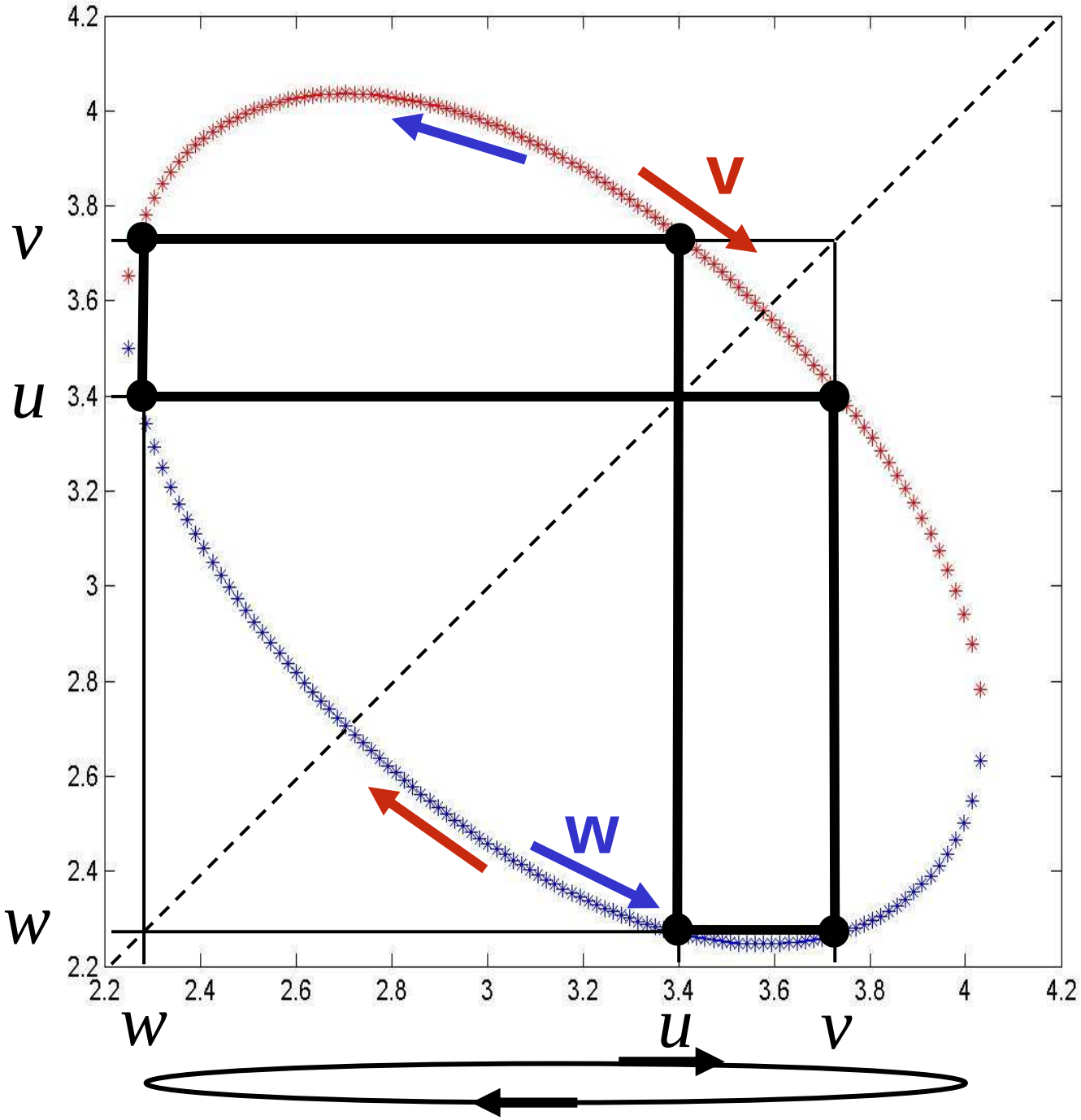
$$\left(A_{22}u_1^2 + A_{02}\right)u_2^2 + A_{11}u_1u_2 + \left(A_{20}u_1^2 + A_{00}\right) = 0$$

$$\left(B_{22}u_2^2 + B_{02}\right)u_3^2 + B_{11}u_2u_3 + \left(B_{20}u_2^2 + B_{00}\right) = 0$$

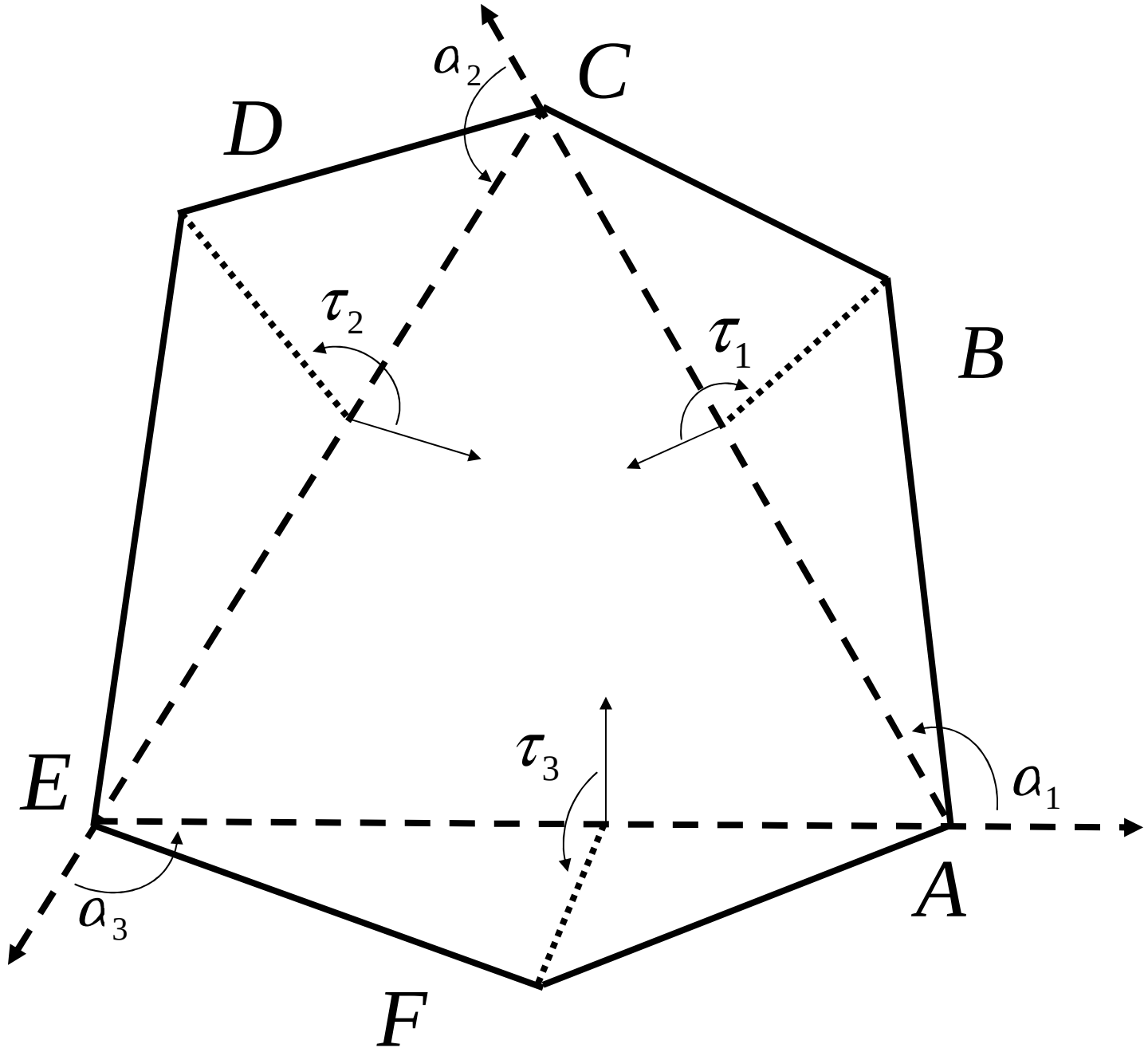
$$\left(C_{22}u_3^2 + C_{02}\right)u_1^2 + C_{11}u_3u_1 + \left(C_{20}u_3^2 + C_{00}\right) = 0$$

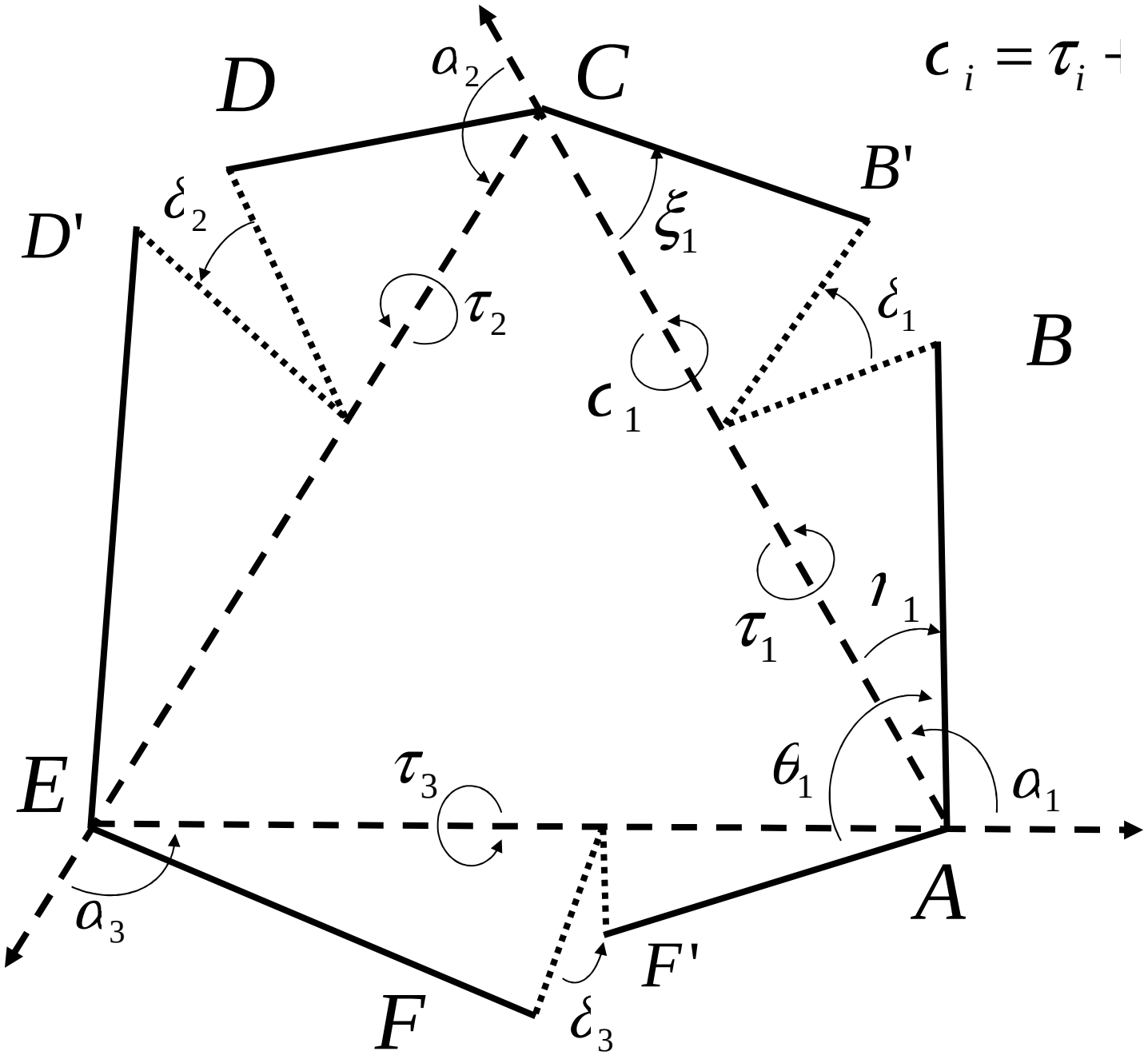
$$u_i = \tan\tau_i / 2$$

General 6-member ring (Octahedron)



conformation circle for cyclohexane twist-boat. As u spans the feasible range, so do v and w. Any triplet (u,v,w) corresponds to a molecular conformation


α_2
C
D
τ_2
τ_1
B
E
τ_3
α_1
A
α_3
F


$\varsigma_i = \tau_i - \delta_i$
D
C
B'
D'
B
E
A
F'
F
α_2
δ_2
ξ_1
τ_2
δ_1
ς_1
τ_1
α_1
θ_1
τ_3
α_3
δ_3

$$\left(A_{22}v_1^2 + A_{02}\right)u_2^2 + A_{11}v_1u_2 + \left(A_{20}v_1^2 + A_{00}\right) = 0$$

$$\left(B_{22}v_2^2 + B_{02}\right)u_3^2 + B_{11}v_2u_3 + \left(B_{20}v_2^2 + B_{00}\right) = 0$$

$$\left(C_{22}v_3^2 + C_{02}\right)u_1^2 + C_{11}v_3u_1 + \left(C_{20}v_3^2 + C_{00}\right) = 0$$

$$v_i = \frac{u_i + \Delta_i}{1 - \Delta_i u_i}, \Delta_i = \tan\delta_i / 2, v_i = \tan\left(\frac{\sigma_i}{2}\right)$$

Three constraint equations:

three generalized angles

$$\left(A_{22}u_2^2 + A_{21}u_2 + A_{20}\right)u_1^2 +$$
$$\left(A_{12}u_2^2 + A_{11}u_2 + A_{10}\right)u_1 +$$
$$\left(A_{02}u_2^2 + A_{01}u_2 + A_{00}\right) = 0$$

$$B_2(u_3)u_2^2 + B_1(u_3)u_2 + B_0(u_3) = 0$$

$$C_2(u_3)u_1^2 + C_1(u_3)u_1 + C_0(u_3) = 0$$

Elimination in Polynomial systems

- Pairwise elimination: Sylvester resultant
- Simultaneous elimination: Dixon resultant
- Both methods lead to a single polynomial in one of the variables: 16^{th} degree
- Differ in complexity and numerical accuracy

Dixon resultant: details

Elimination matrix

$$C(u_1,u_2,v_1,v_2)=\begin{bmatrix} p_1(u_1,u_2) & p_2(u_2) & p_3(u_1) \\ p_1(v_1,u_2) & p_2(u_2) & p_3(v_1) \\ p_1(v_1,v_2) & p_2(v_2) & p_3(v_1) \end{bmatrix}$$

Dixon polynomial

$$d(u_1,u_2,v_1,v_2)=\frac{\det C}{(u_1-v_1)(u_2-v_2)}$$

Vanishes at all common roots of original system.

Dixon matrix: coefficients of Dixon polynomial, arranged by monomials in x, y

$$d(u_1, u_2, v_1, v_2) = VDU$$

$$V = \begin{bmatrix} 1 & v_1 & v_2 & v_1 v_2 & v_2^2 & v_1 v_2^2 & v_2^3 & v_1 v_2^3 \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & u_1 & u_1^2 & u_1^3 & u_2 & u_1 u_2 & u_1^2 u_2 & u_1^3 u_2 \end{bmatrix}^T$$

If u1, u2 are roots of original system, U becomes a right null vector. Therefore the vanishing of det D is a necessary condition for the existence of a common root.

The Dixon resultant contains an extraneous factor (deg. 16)

Coutsias, Seok, Wester, Dill, JCC 2005

$$D = (A_{22} B_2 C_2)^4 \det S$$

$$S :=$$

$$\begin{bmatrix}
0 & \begin{vmatrix} B_0 & B_1 \\ A_{00} & A_{01} \end{vmatrix} & \begin{vmatrix} B_0 & B_1 \\ A_{10} & A_{11} \end{vmatrix} & \begin{vmatrix} B_0 & B_1 \\ A_{20} & A_{21} \end{vmatrix} & 0 & \begin{vmatrix} B_0 & B_2 \\ A_{00} & A_{02} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{10} & A_{12} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{20} & A_{22} \end{vmatrix} \\
\begin{vmatrix} B_0 & B_1 \\ A_{00} & A_{01} \end{vmatrix} & \begin{vmatrix} B_0 & B_1 \\ A_{10} & A_{11} \end{vmatrix} & \begin{vmatrix} B_0 & B_1 \\ A_{20} & A_{21} \end{vmatrix} & 0 & \begin{vmatrix} B_0 & B_2 \\ A_{00} & A_{02} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{10} & A_{12} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{20} & A_{22} \end{vmatrix} & 0 \\
0 & \begin{vmatrix} B_0 & B_2 \\ A_{00} & A_{02} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{10} & A_{12} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{20} & A_{22} \end{vmatrix} & 0 & \begin{vmatrix} B_1 & B_2 \\ A_{01} & A_{02} \end{vmatrix} & \begin{vmatrix} B_1 & B_2 \\ A_{11} & A_{12} \end{vmatrix} & \begin{vmatrix} B_1 & B_2 \\ A_{21} & A_{22} \end{vmatrix} \\
\begin{vmatrix} B_0 & B_2 \\ A_{00} & A_{02} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{10} & A_{12} \end{vmatrix} & \begin{vmatrix} B_0 & B_2 \\ A_{20} & A_{22} \end{vmatrix} & 0 & \begin{vmatrix} B_1 & B_2 \\ A_{01} & A_{02} \end{vmatrix} & \begin{vmatrix} B_1 & B_2 \\ A_{11} & A_{12} \end{vmatrix} & \begin{vmatrix} B_1 & B_2 \\ A_{21} & A_{22} \end{vmatrix} & 0 \\
0 & 0 & 0 & 0 & 0 & C_0 & C_1 & C_2 \\
0 & 0 & 0 & 0 & C_0 & C_1 & C_2 & 0 \\
0 & C_0 & C_1 & C_2 & 0 & 0 & 0 & 0 \\
C_0 & C_1 & C_2 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

$$S_i = \begin{bmatrix} 0 & L_{0i} & L_{1i} & L_{2i} & 0 & M_{0i} & M_{1i} & M_{2i} \\ L_{0i} & L_{1i} & L_{2i} & 0 & M_{0i} & M_{1i} & M_{2i} & 0 \\ 0 & M_{0i} & M_{1i} & M_{2i} & 0 & N_{0i} & N_{1i} & N_{2i} \\ M_{0i} & M_{1i} & M_{2i} & 0 & N_{0i} & N_{1i} & N_{2i} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{0i} & C_{1i} & C_{2i} \\ 0 & 0 & 0 & 0 & C_{0i} & C_{1i} & C_{2i} & 0 \\ 0 & C_{0i} & C_{1i} & C_{2i} & 0 & 0 & 0 & 0 \\ C_{0i} & C_{1i} & C_{2i} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$L_k = L_{k2} u_3^2 + L_{k1} u_3 + L_{k0}$$

$$S = S_2 u_3^2 + S_1 u_3 + S_0$$

Eliminating the extraneous factor there results a matrix polynomial of degree 2; its determinant can be computed with a companion matrix of size 16X16. Then the real eigenvalues give u3 for all possible conformations. Values for the other two variables are found from the appropriate eigenvector components

$$\det\left(\begin{bmatrix} I & 0 \\ 0 & S_3 \end{bmatrix} u_3 - \begin{bmatrix} 0 & I \\ -S_0 & -S_1 \end{bmatrix}\right) = 0$$

Successive elimination (Sylvester resultant method) results in a matrix of size 6X6 but with quartic coefficients, leading to a companion matrix of size 24X24 (but still a 16^{th} degree polynomial)

QZ or Characteristic Polynomial

Either:

- (1) Solve generalized eigenproblem using **QZ** algorithm (cubic order in matrix size, here approx. 16^3 ~4kflops)

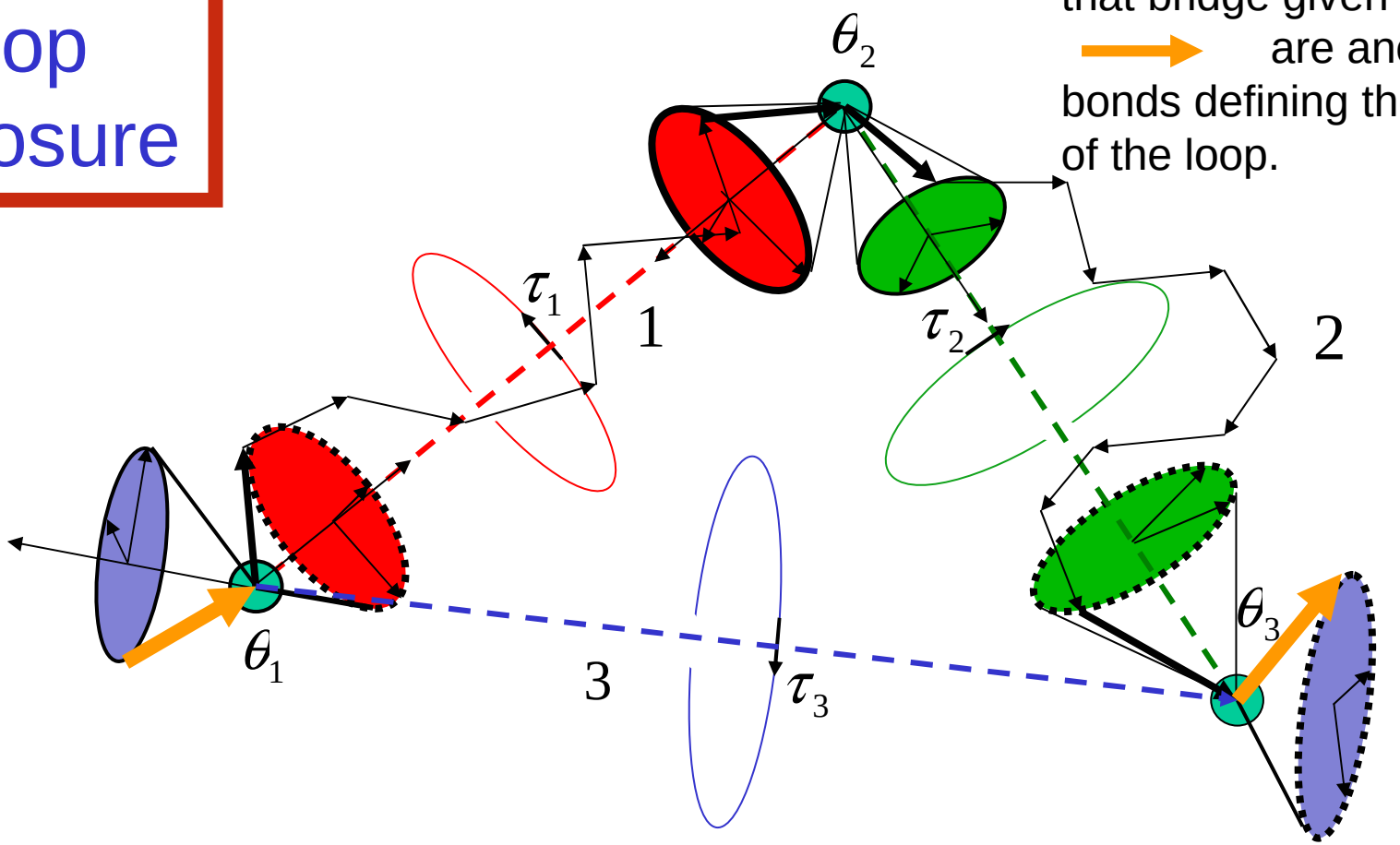
Or:

- (2a) Use **Lagrange's expansion** in complementary minors to efficiently compute the characteristic polynomial of the Dixon resultant (~2.2kflops).
- (2b) Use **Sturm sequences** (count number of real zeros between two values) for an efficient computation of real zeros from characteristic polynomial (bisection/Newton: variable, but low, cost)

Triaxial Loop Closure

Method calculates all (0-16) possible loop configurations that bridge given gap.

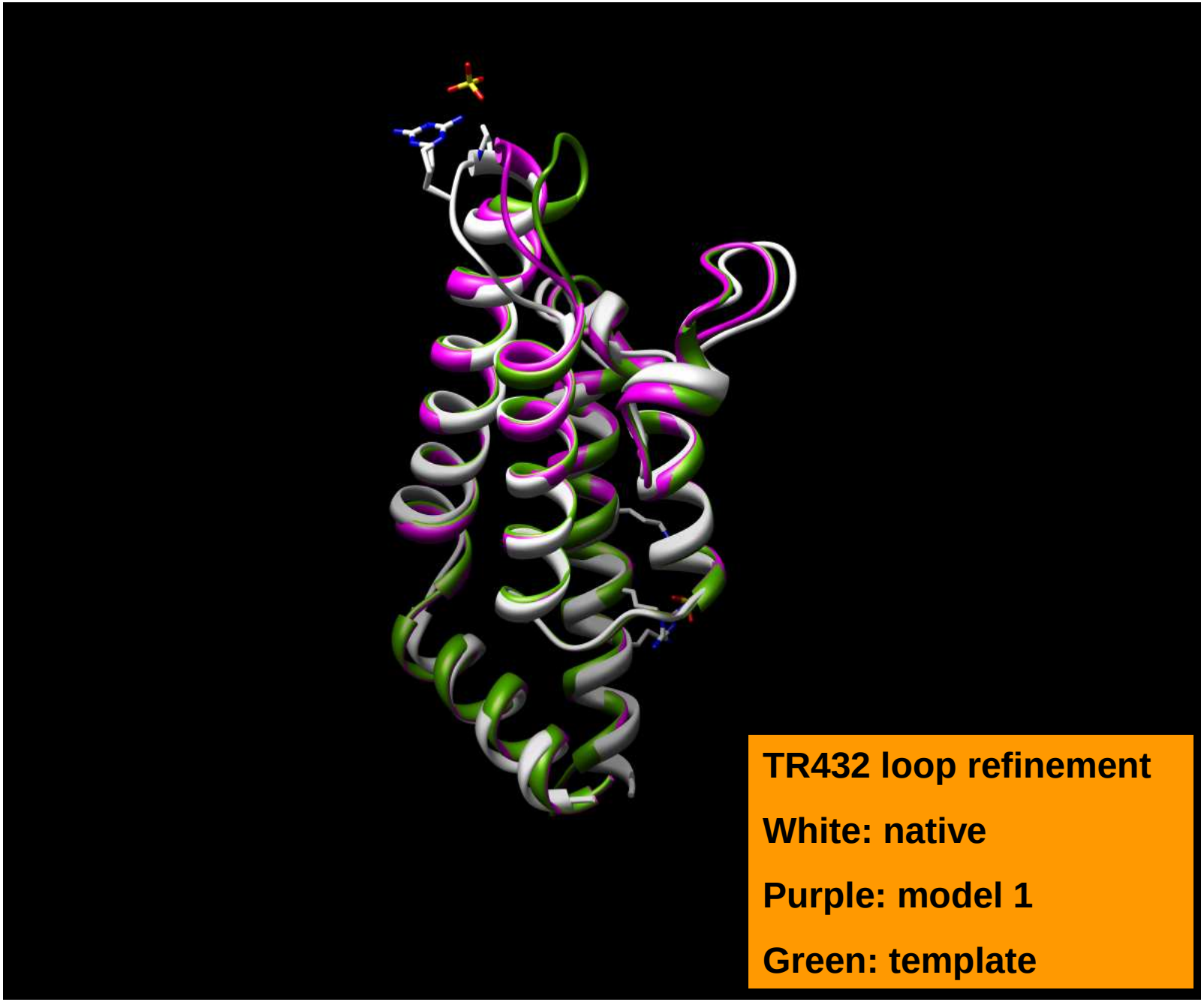
→ are anchor bonds defining the two ends of the loop.

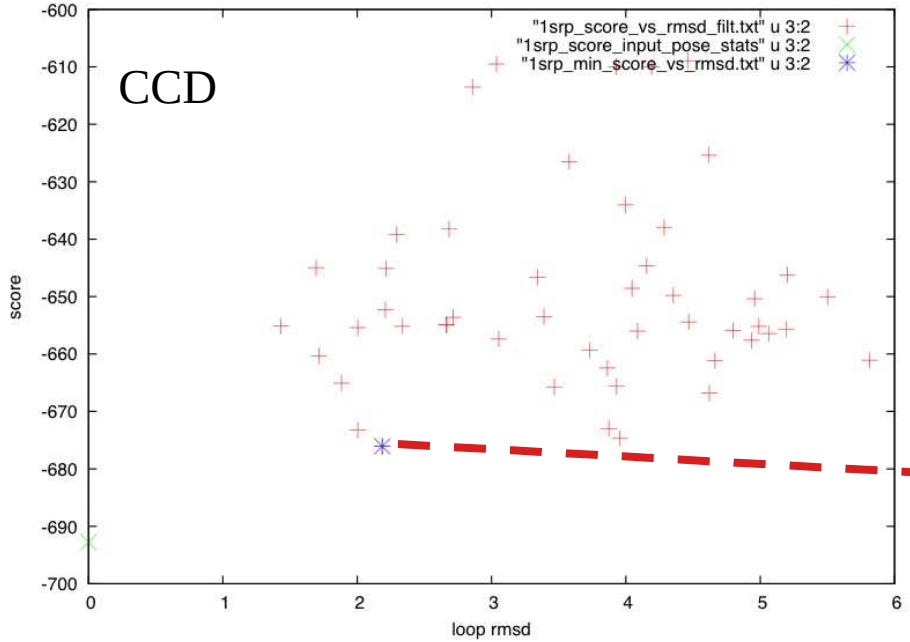


Rotate segments 1,2 about resp. virtual axes by angles τ_1, τ_2

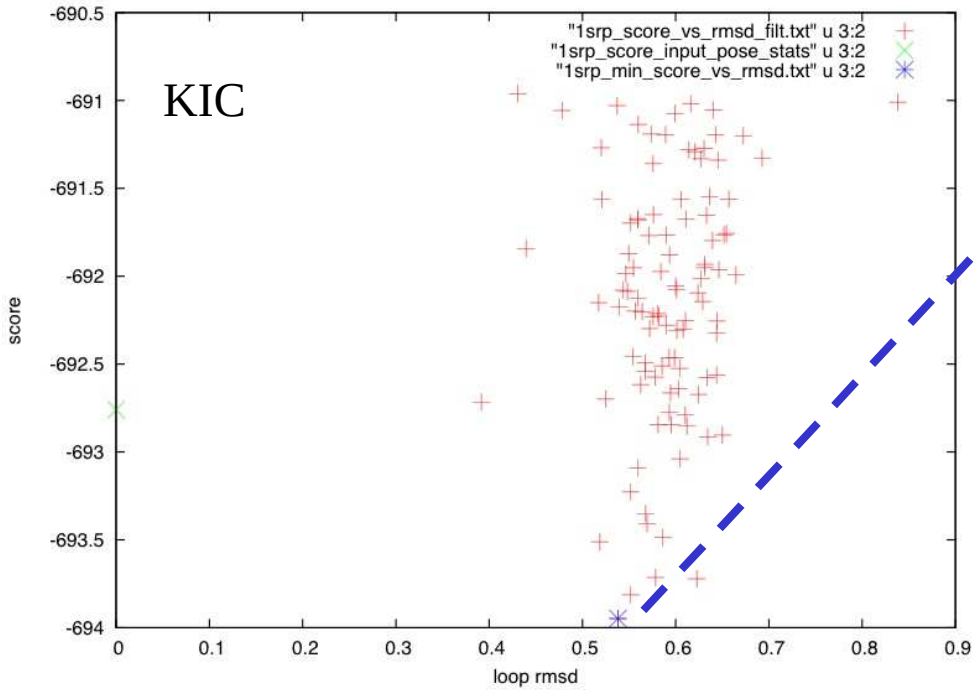
Rotate entire loop about virtual axis 3 by angle τ_3

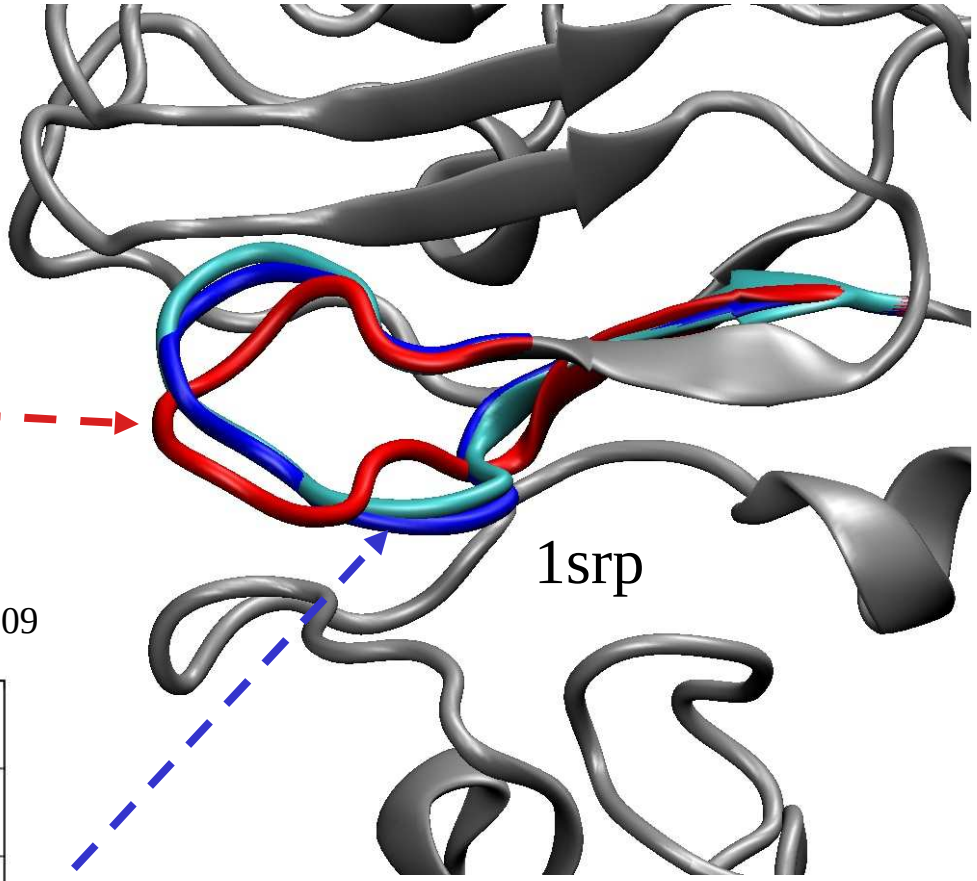
Choose values of the τ_1, τ_2, τ_3 to fix bond angles θ


TR432 loop refinement
White: native
Purple: model 1
Green: template

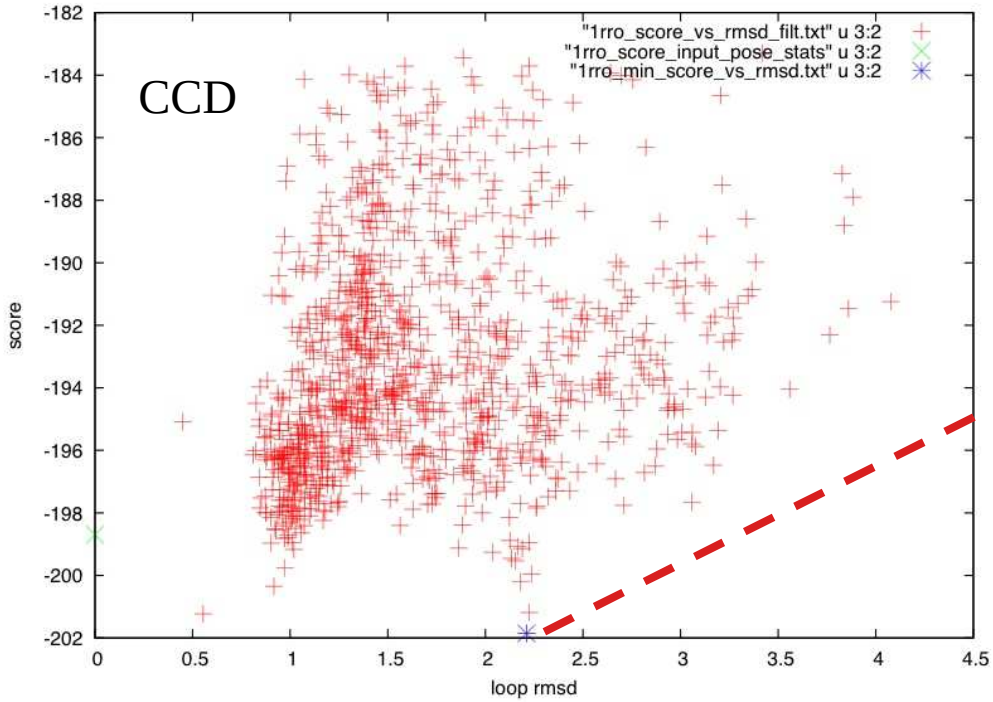


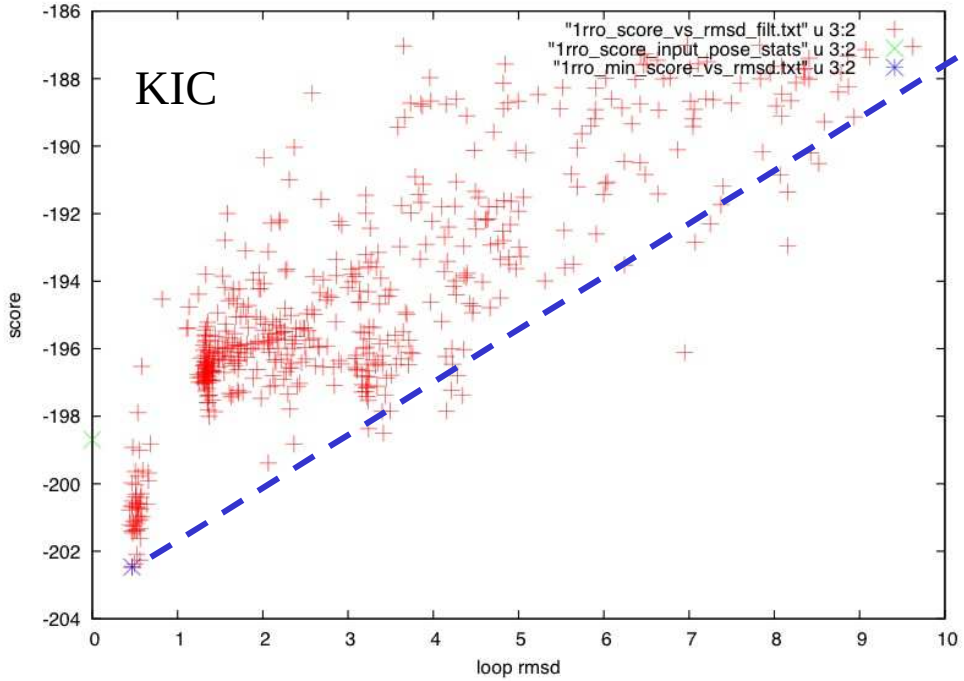
Mandell, Coutsias, Kortemme; Nature Meth. 6(8), 551-2, 2009

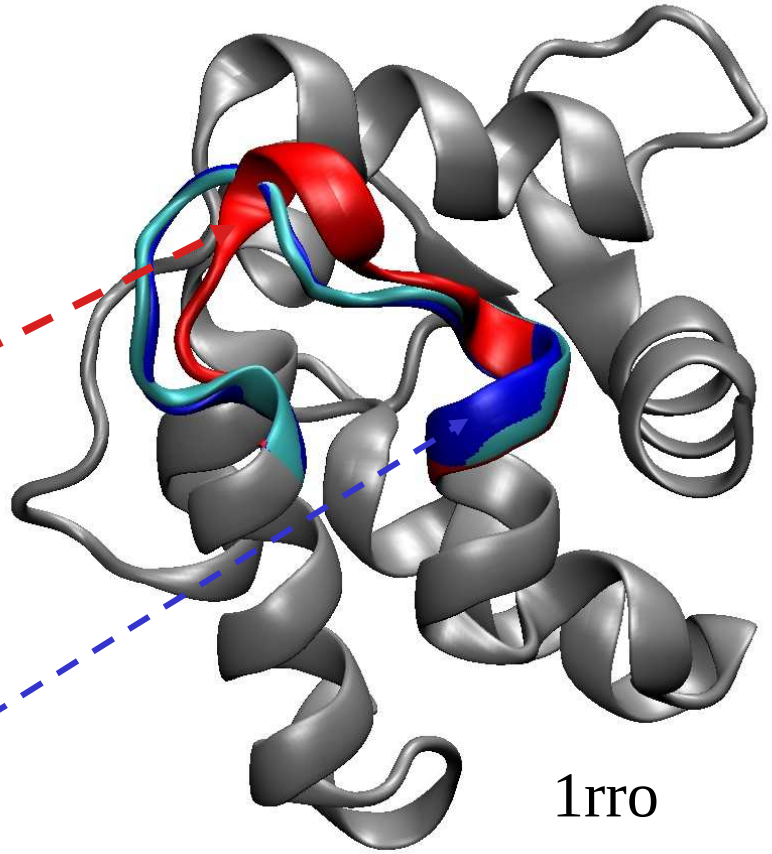




KIC provides superior sampling, compared to current Rosetta protocol (based on Cyclic Coordinate Descent-Canutescu & Dunbrack, JCC 2003)

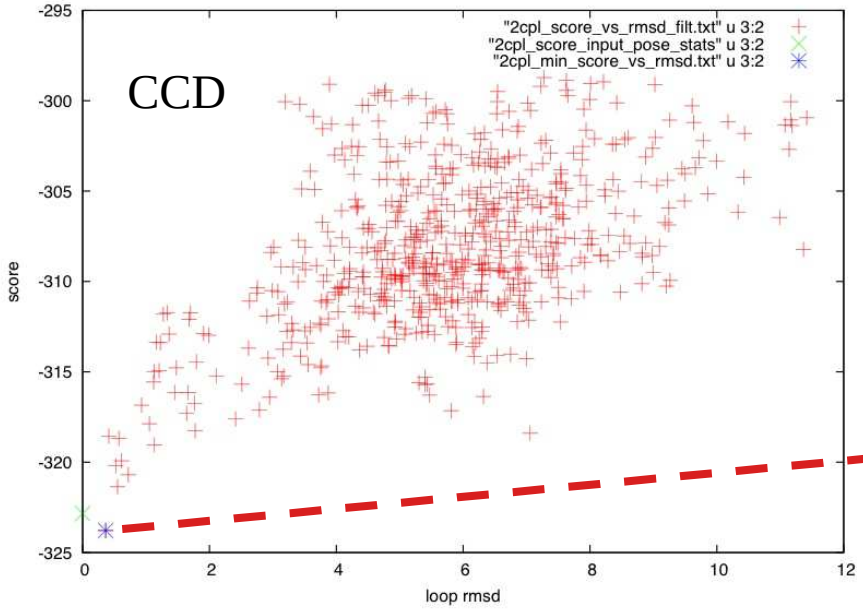


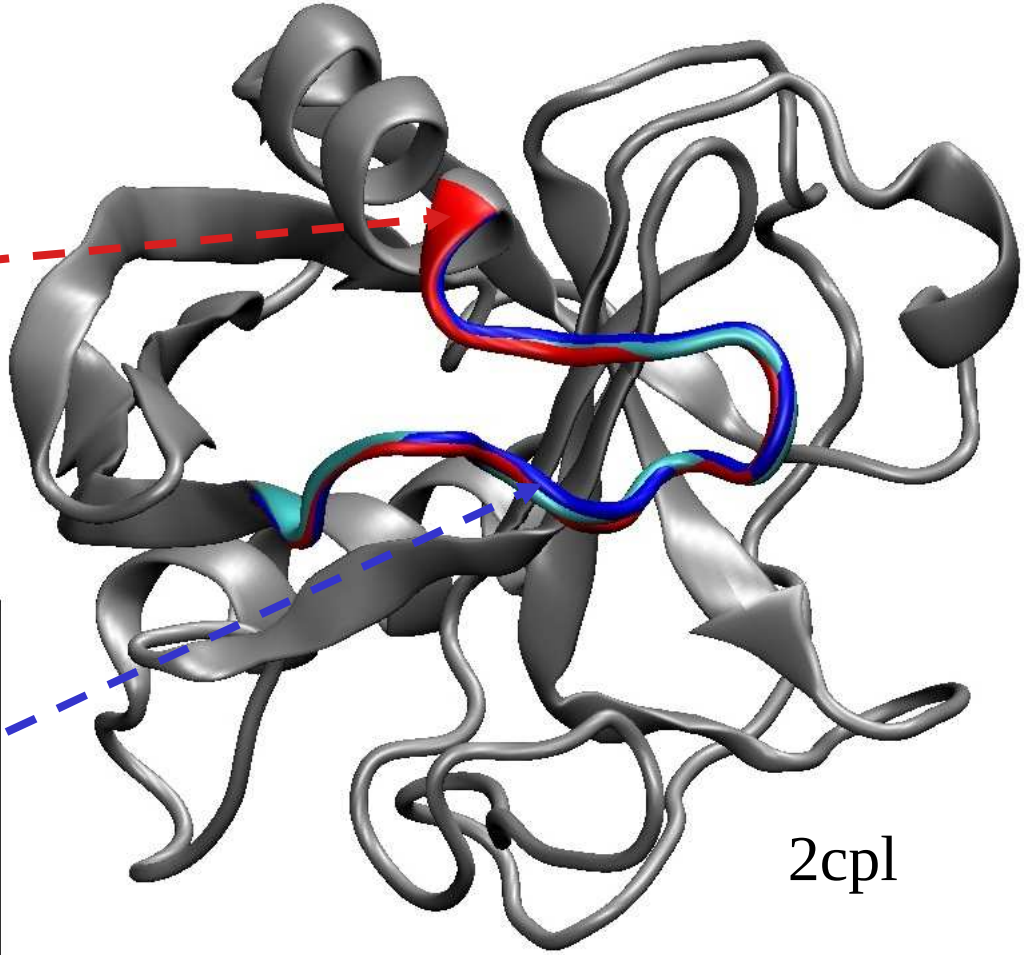


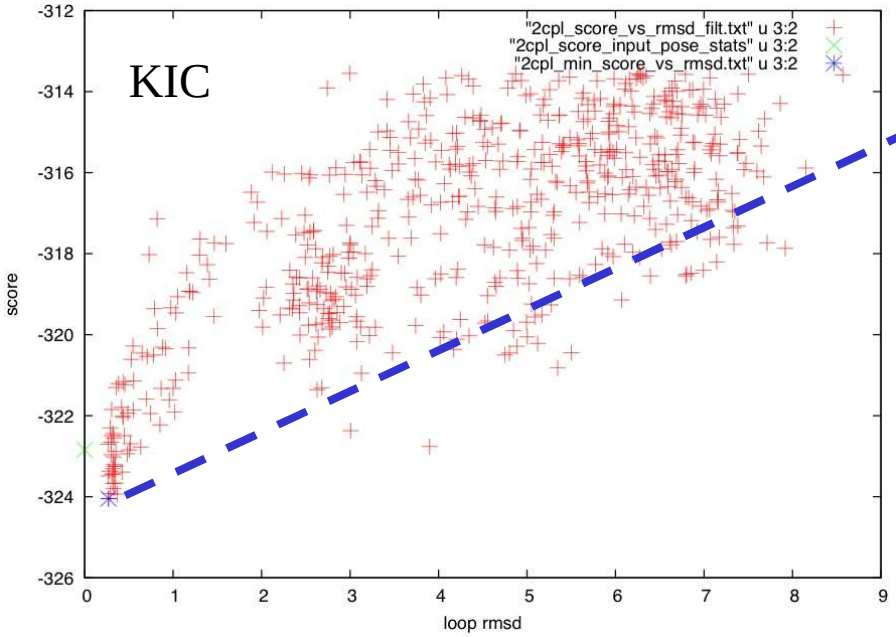


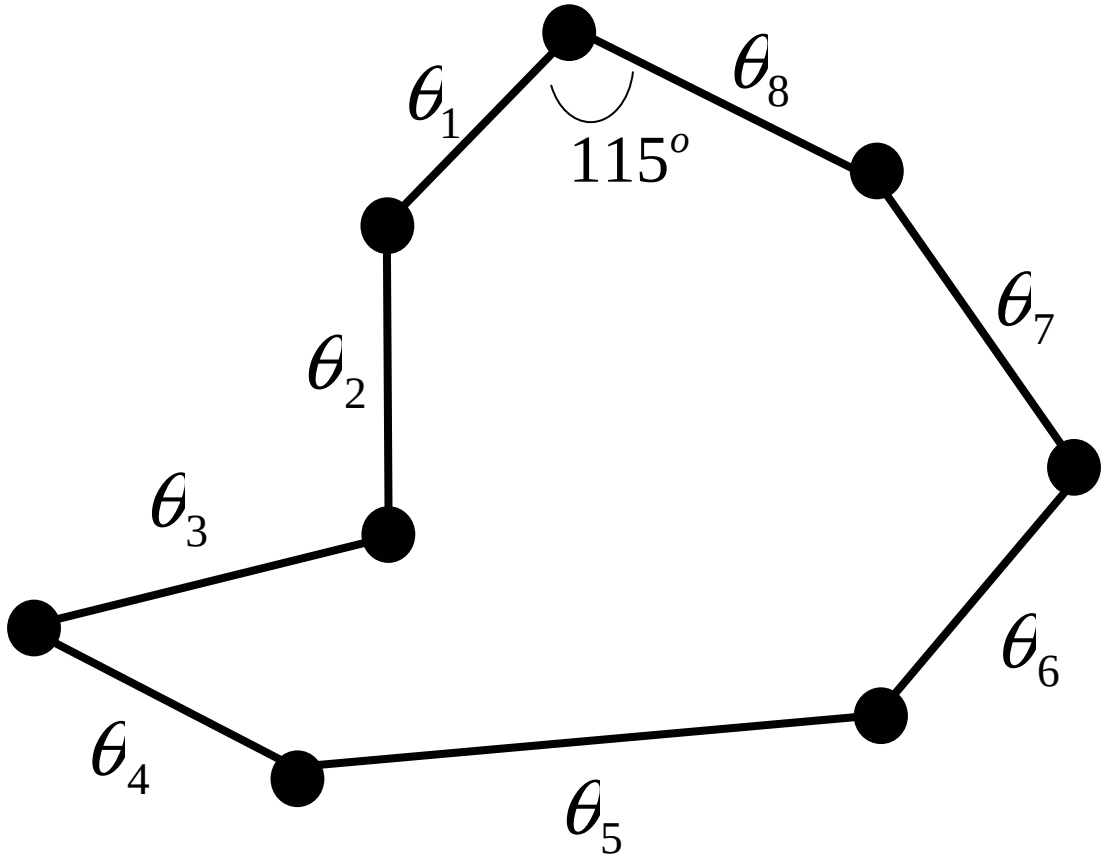
Mandell, Coutsias, Kortemme;
Nature Meth. 6(8), 551-2, 2009

Mandell, Coutsias, Kortemme; Nature Meth. 6(8), 551-2, 2009

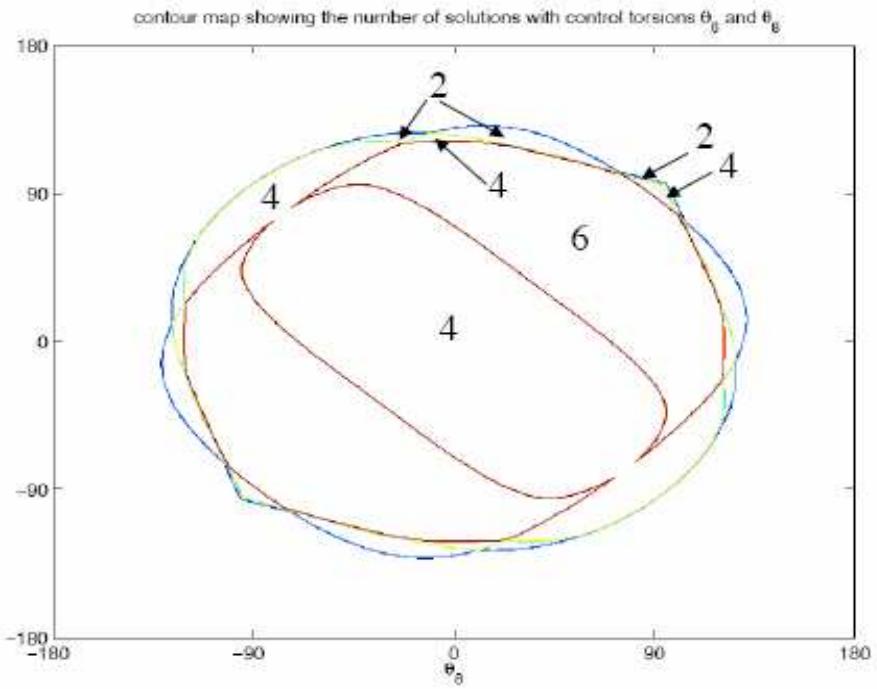


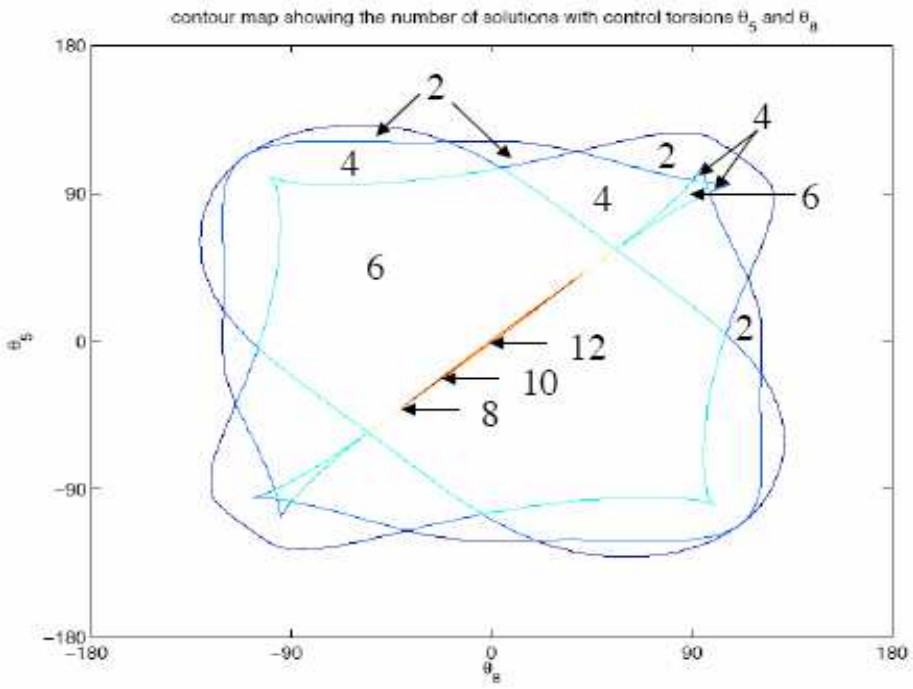


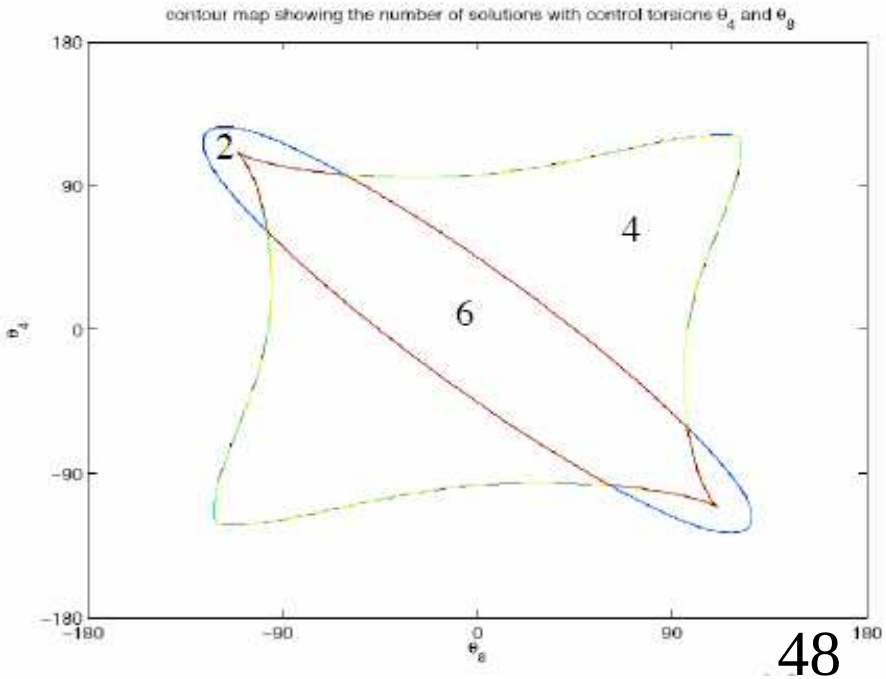




Comparison of algorithms: exhaustive covering of the conformational space of cyclooctane. Two torsions set to arbitrary values, other 6 determined to satisfy closure



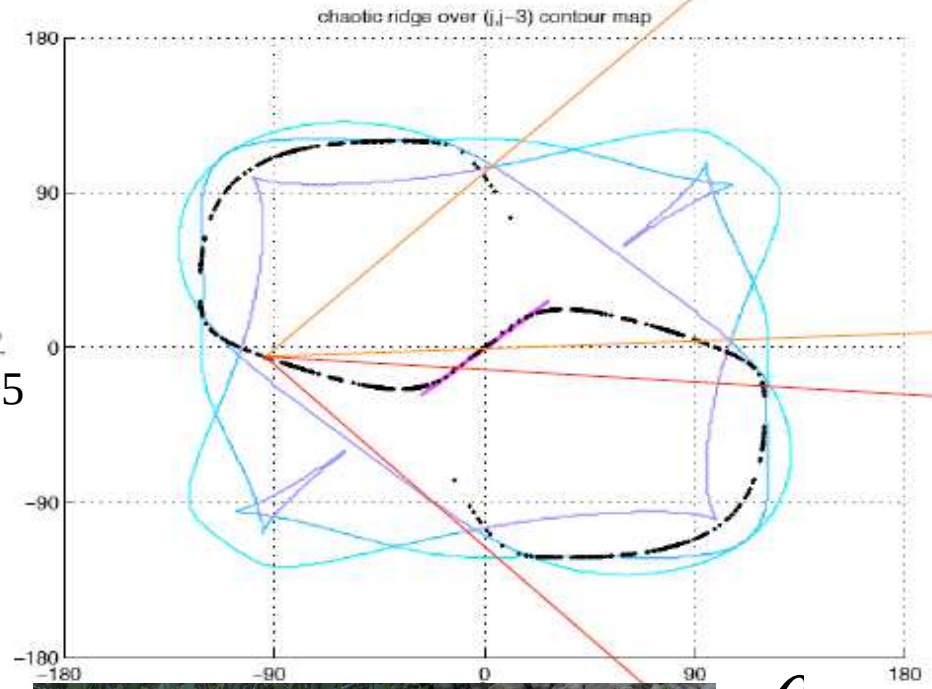


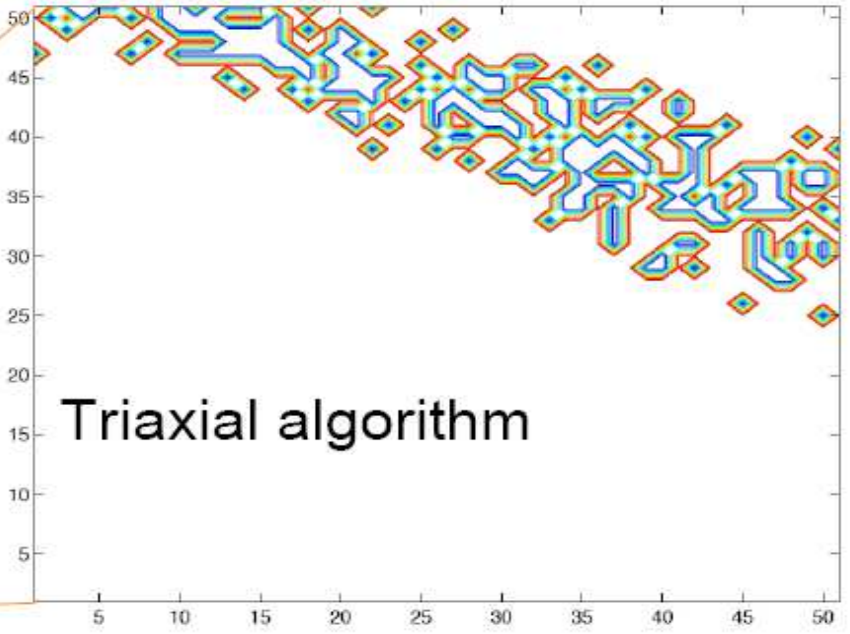


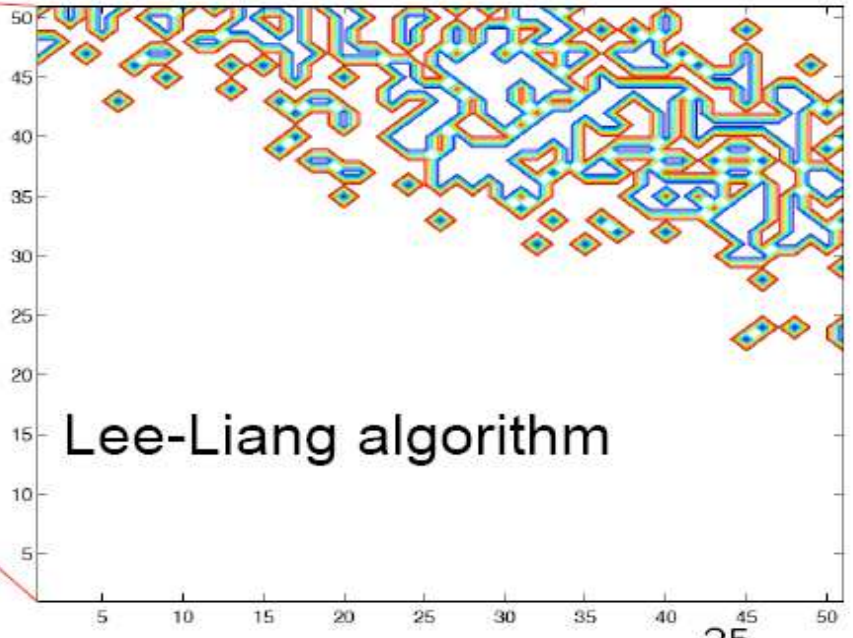
contour maps of the number of solutions (conformations) of cyclooctane

Pollock & Coutsias, (preprint, 2009)

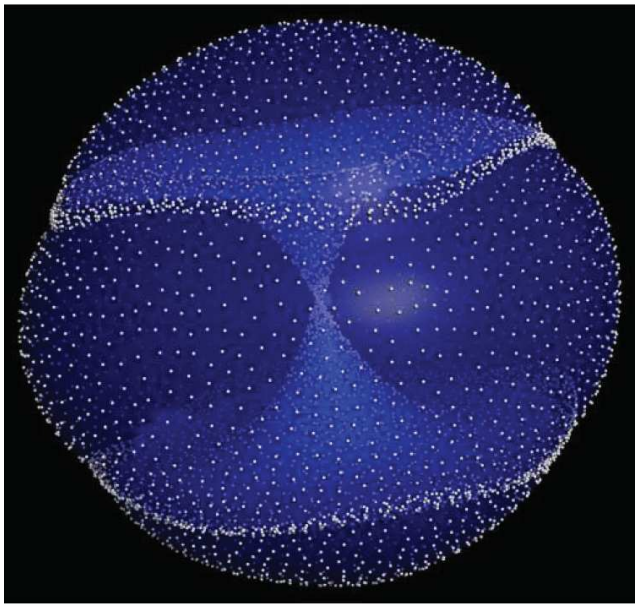
1st chaotic ridge, shown in t5-t8 projection, is a pathology for both TLC and LL algorithms

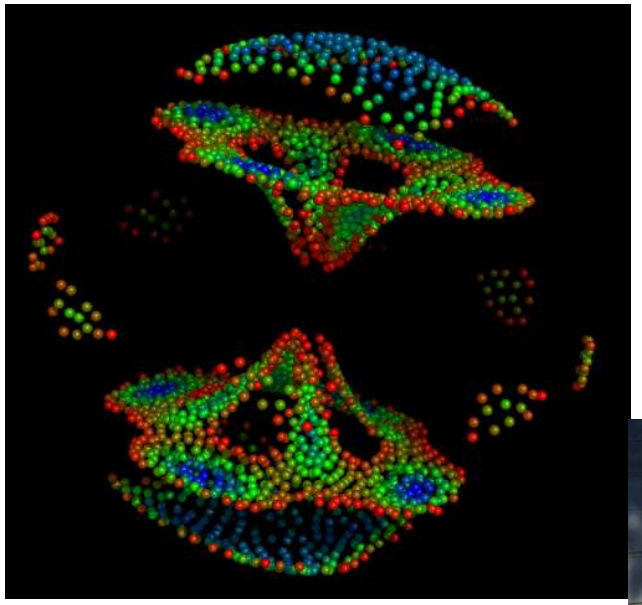


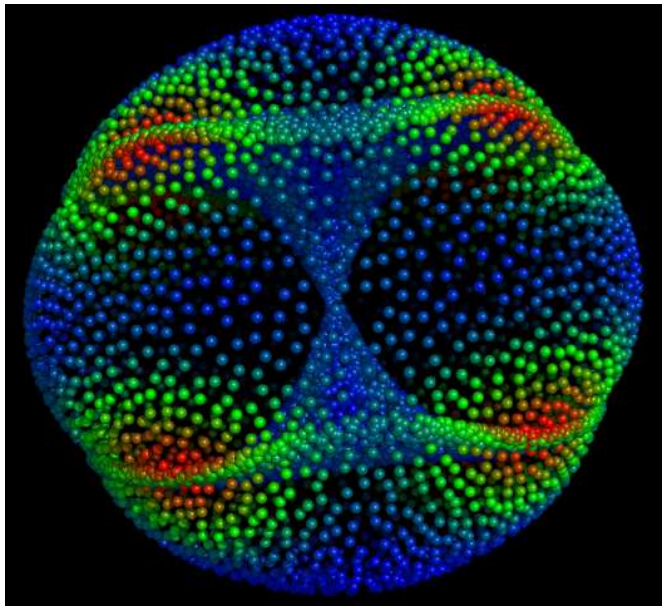


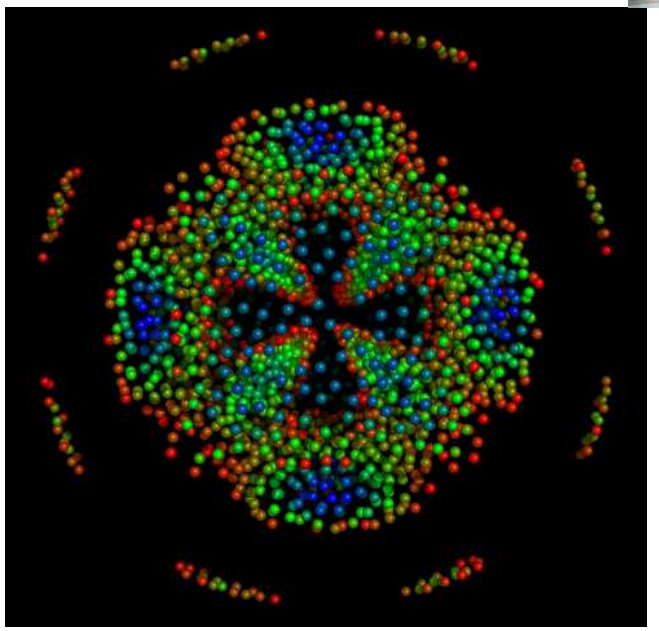


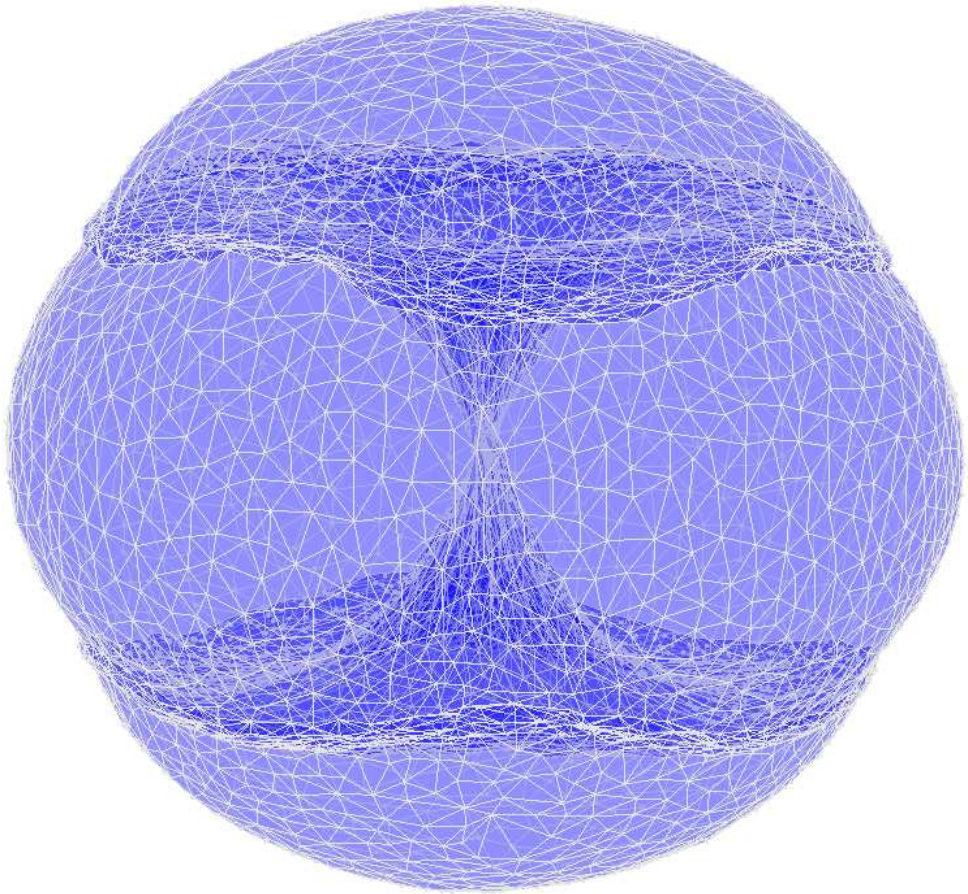
3-D Embedding from Isomap

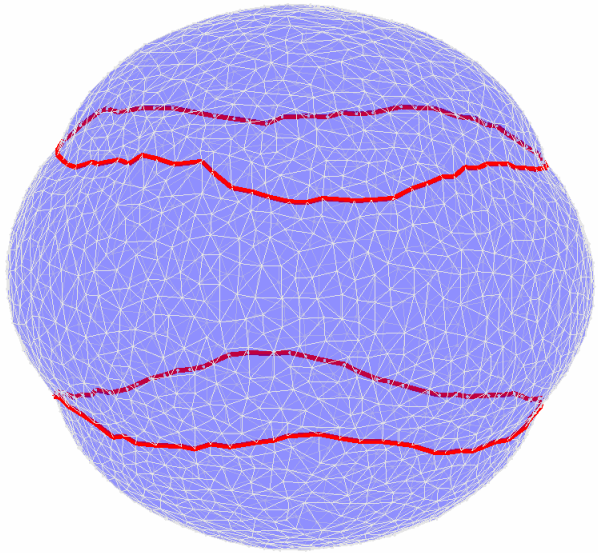


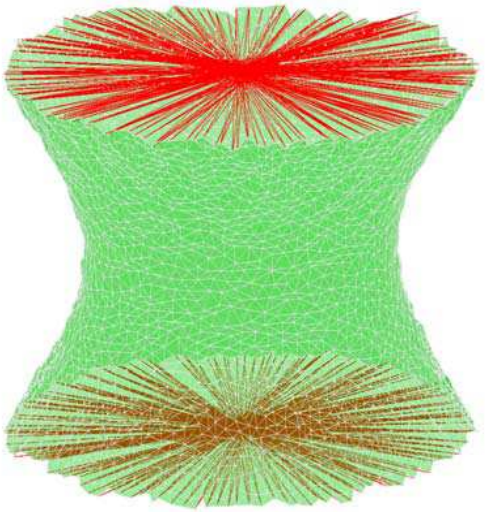




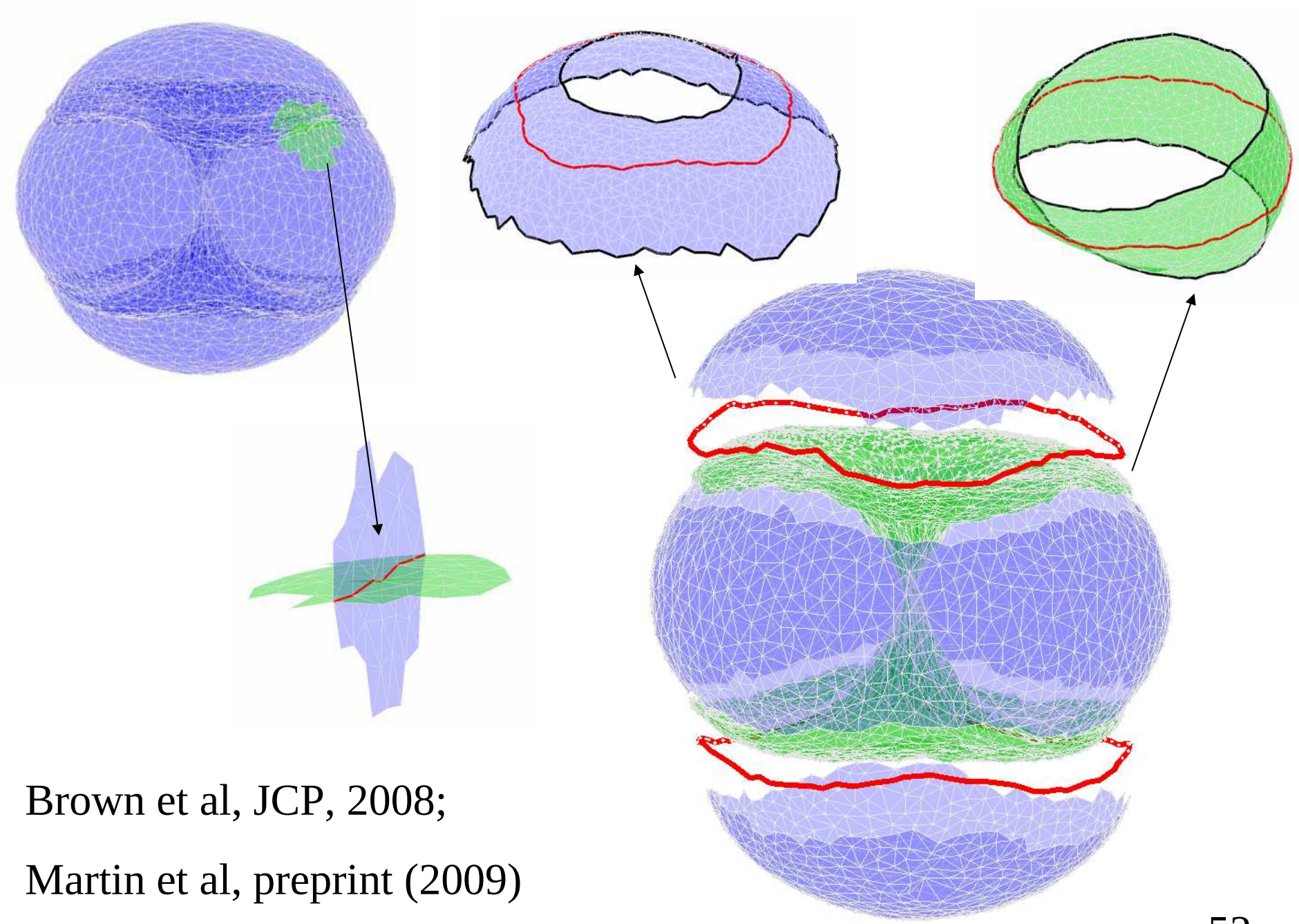








Conformation space of canonical cyclooctane, the simplest 2DoF closed loop: complete cover (.5deg) involving ~1M points, was reduced to high-res cover of 3K pts with ISOMAP; space is an algebraic variety, not a manifold!



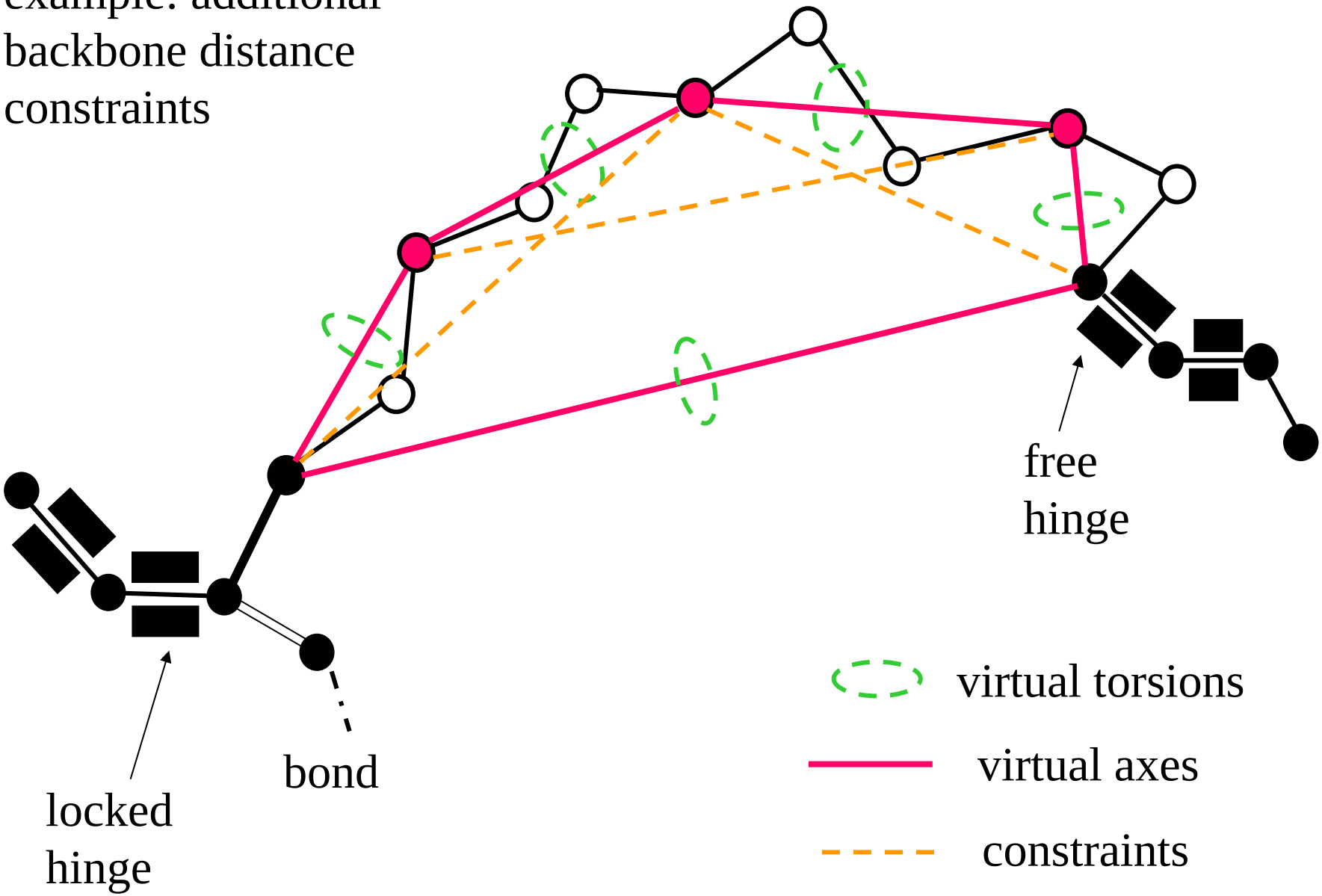
Brown et al, JCP, 2008;

Martin et al, preprint (2009)

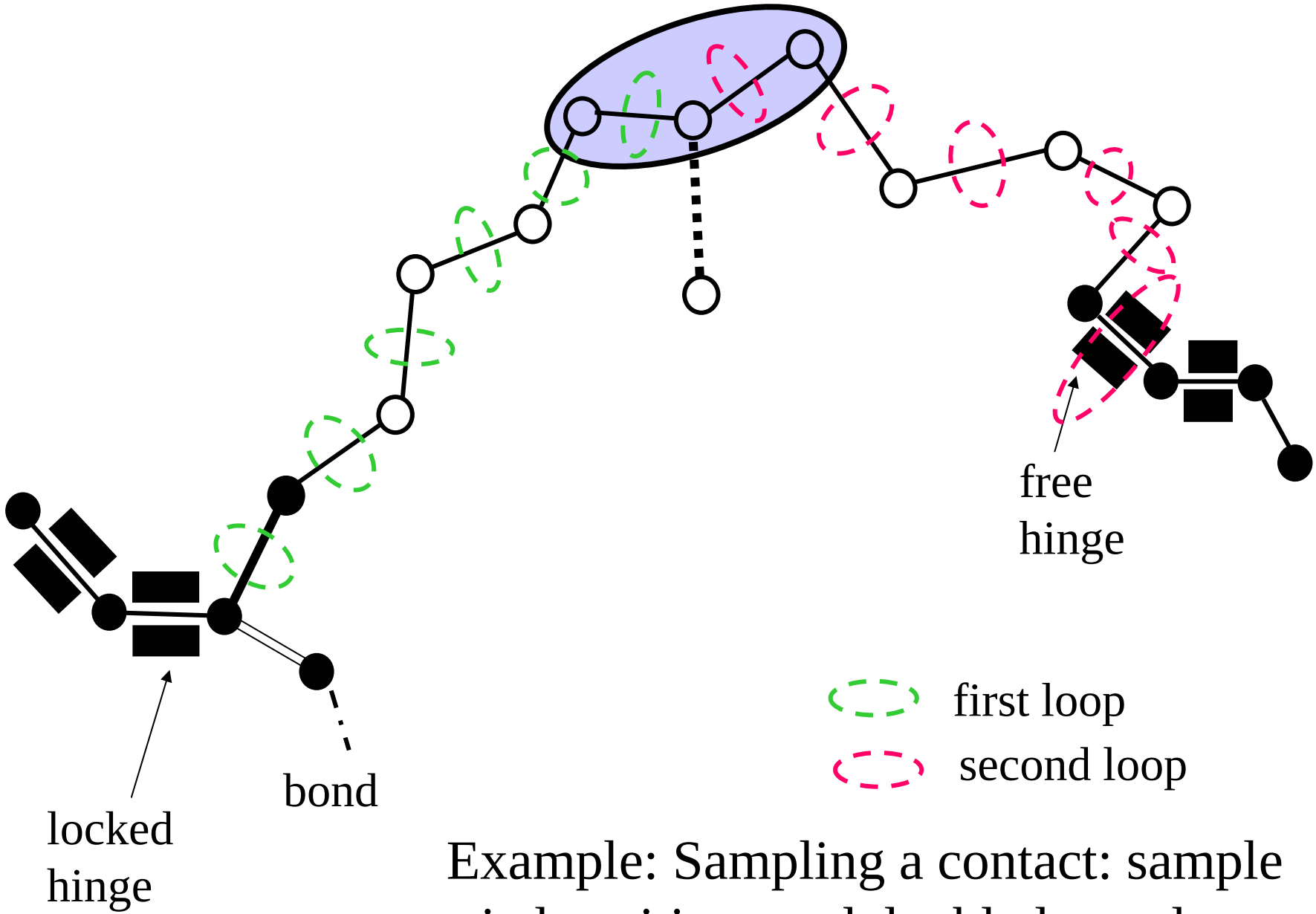
Constrained sampling

- Multiple (long) loops
- Tetrapeptide+distance(s), orientation
- Pentapeptide + localization
- Cysteine bridges, multiple loops
- Uncertain Ca positions
- Sampling a binding pocket

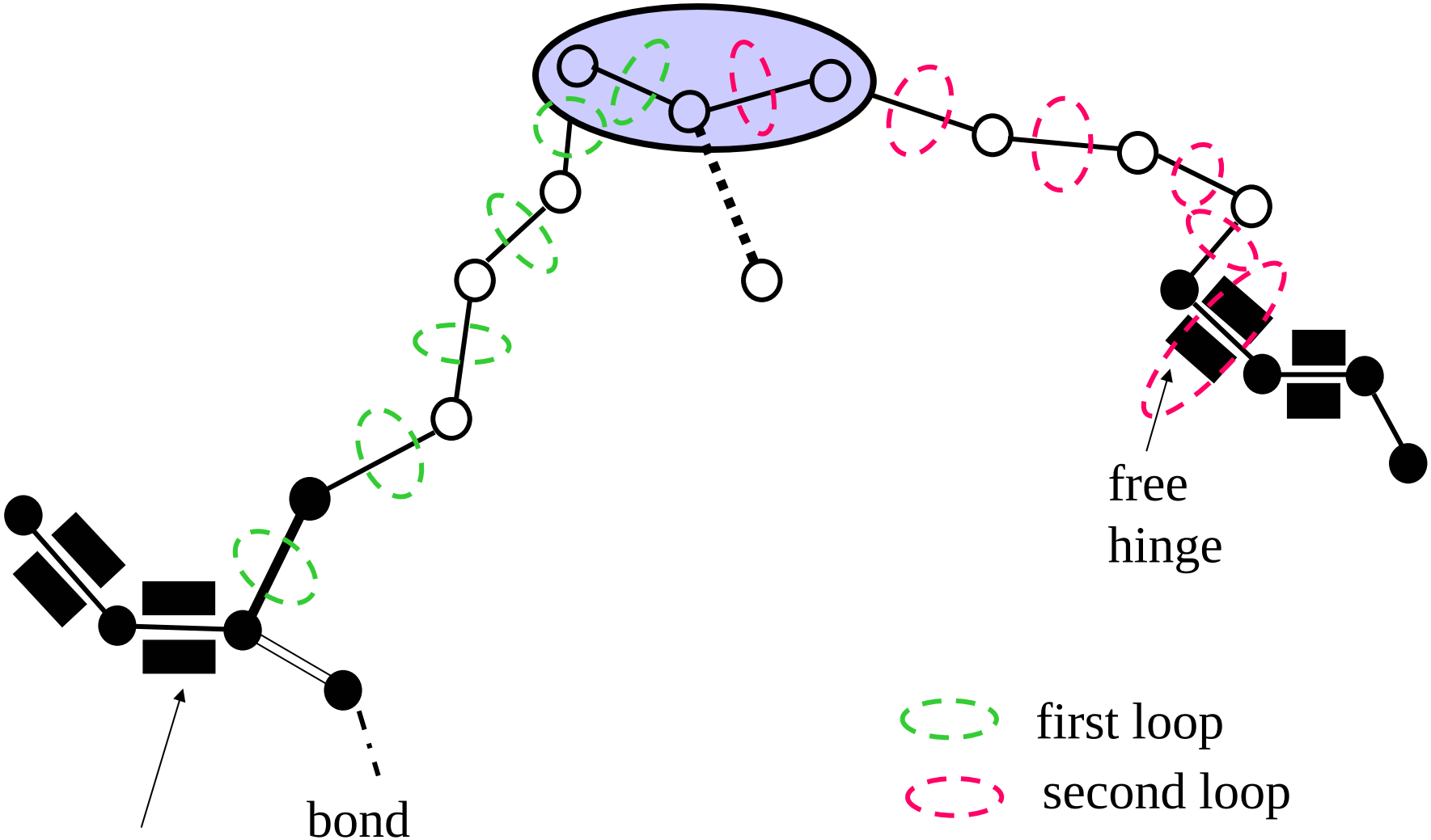
example: additional backbone distance constraints



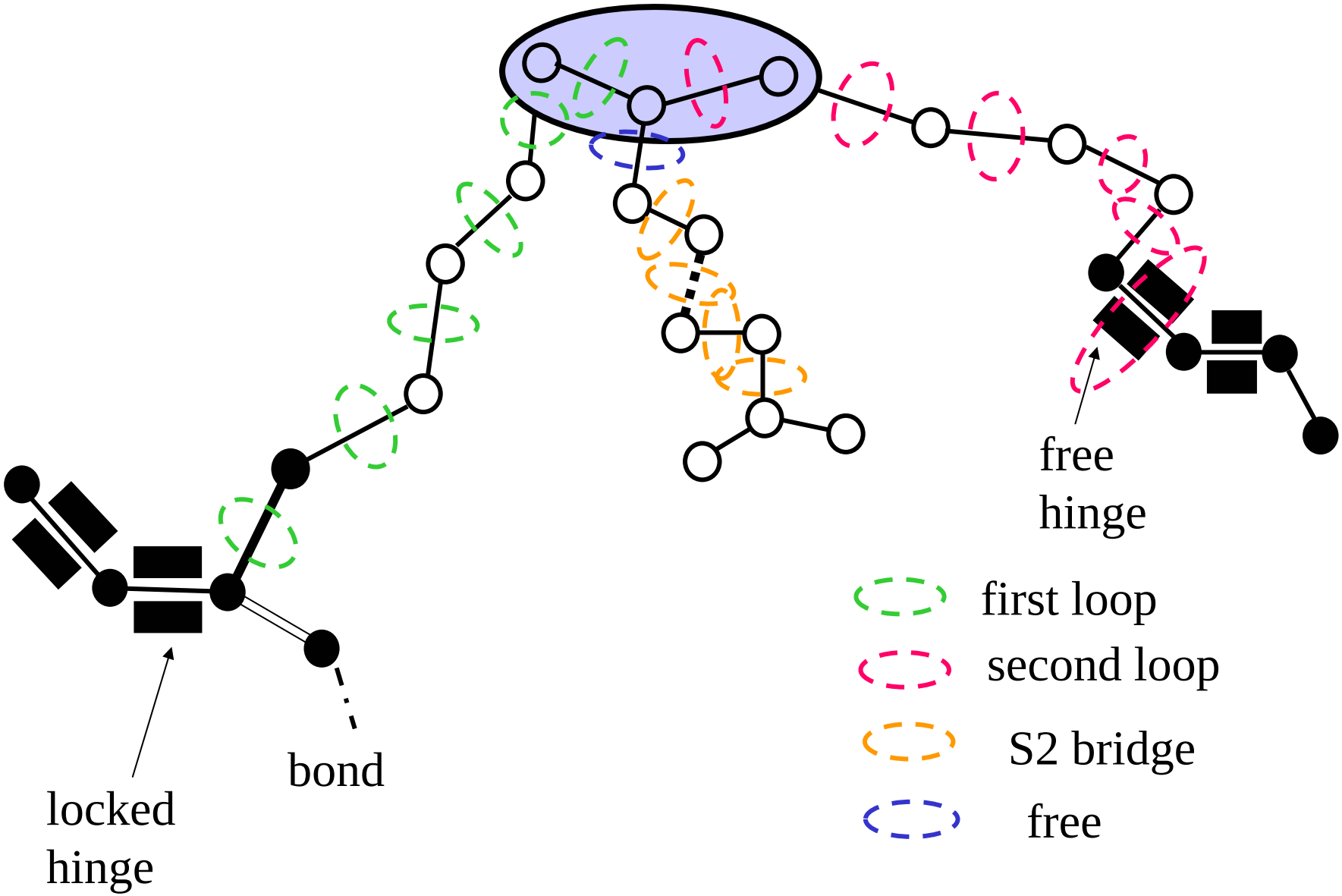
5 coupled tetrahedrals=> 2*2^5 = 64 possible solutions



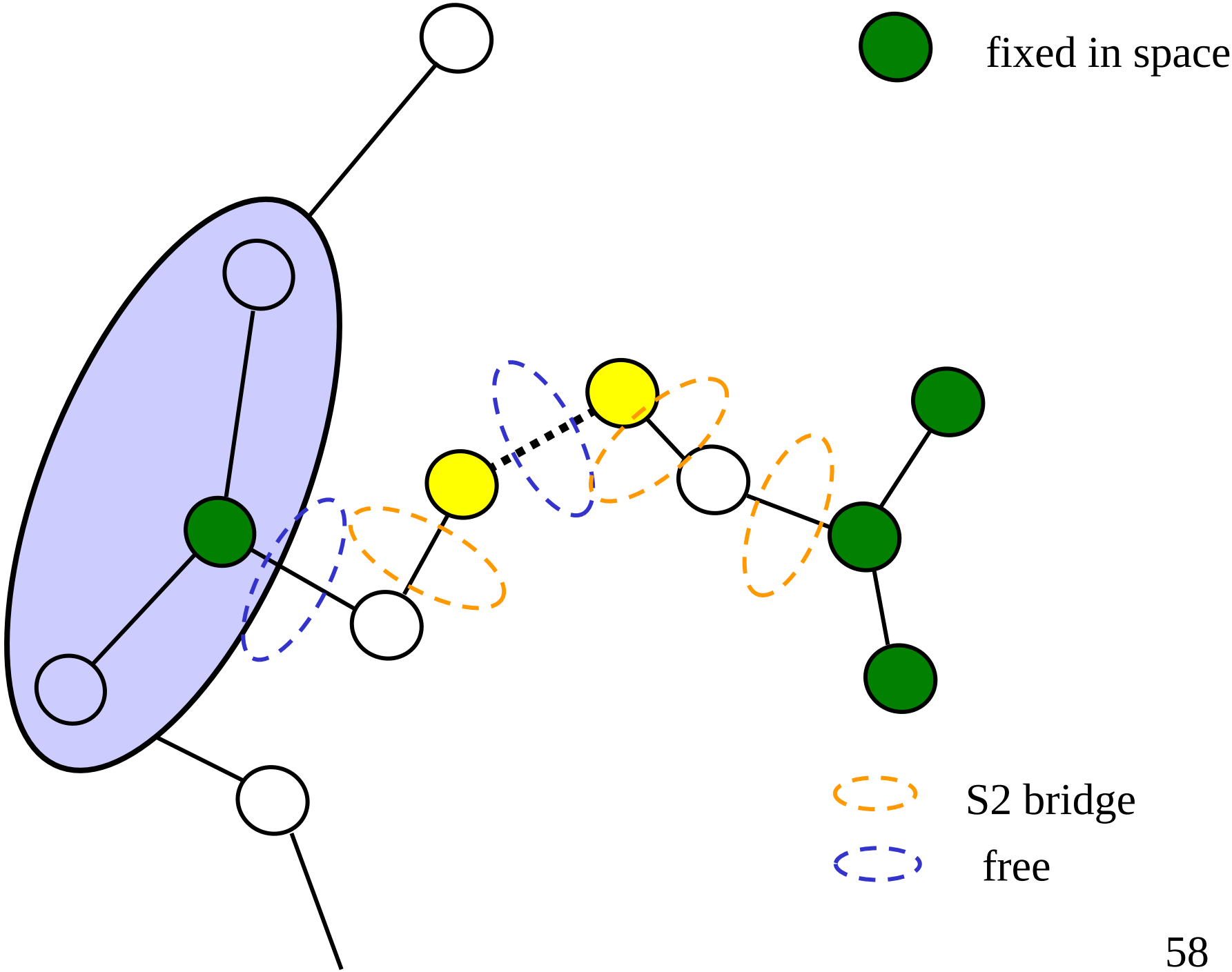
Example: Sampling a contact: sample triad positions and double loop closure

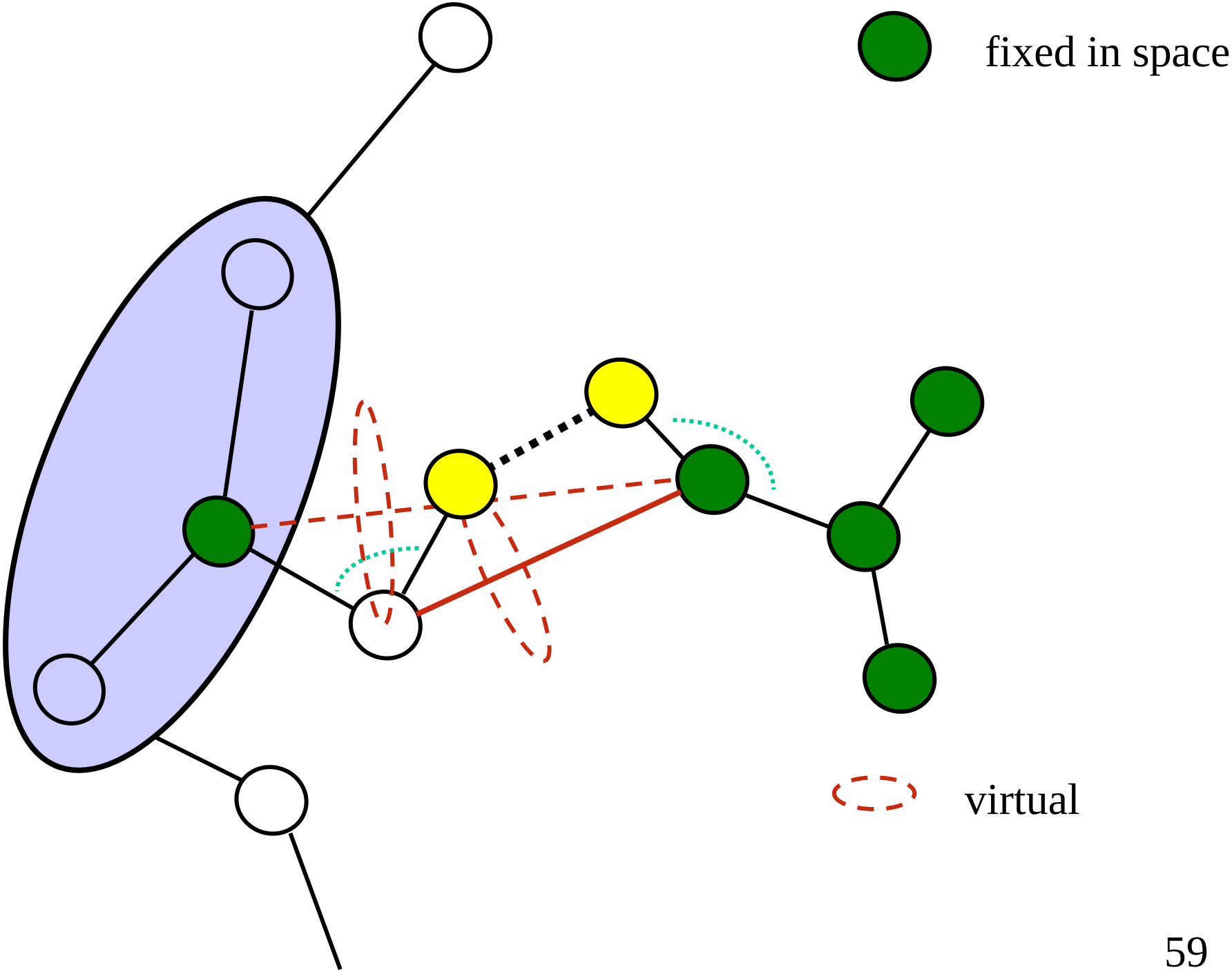


Example: Sampling a contact: sample triad positions and double loop closure

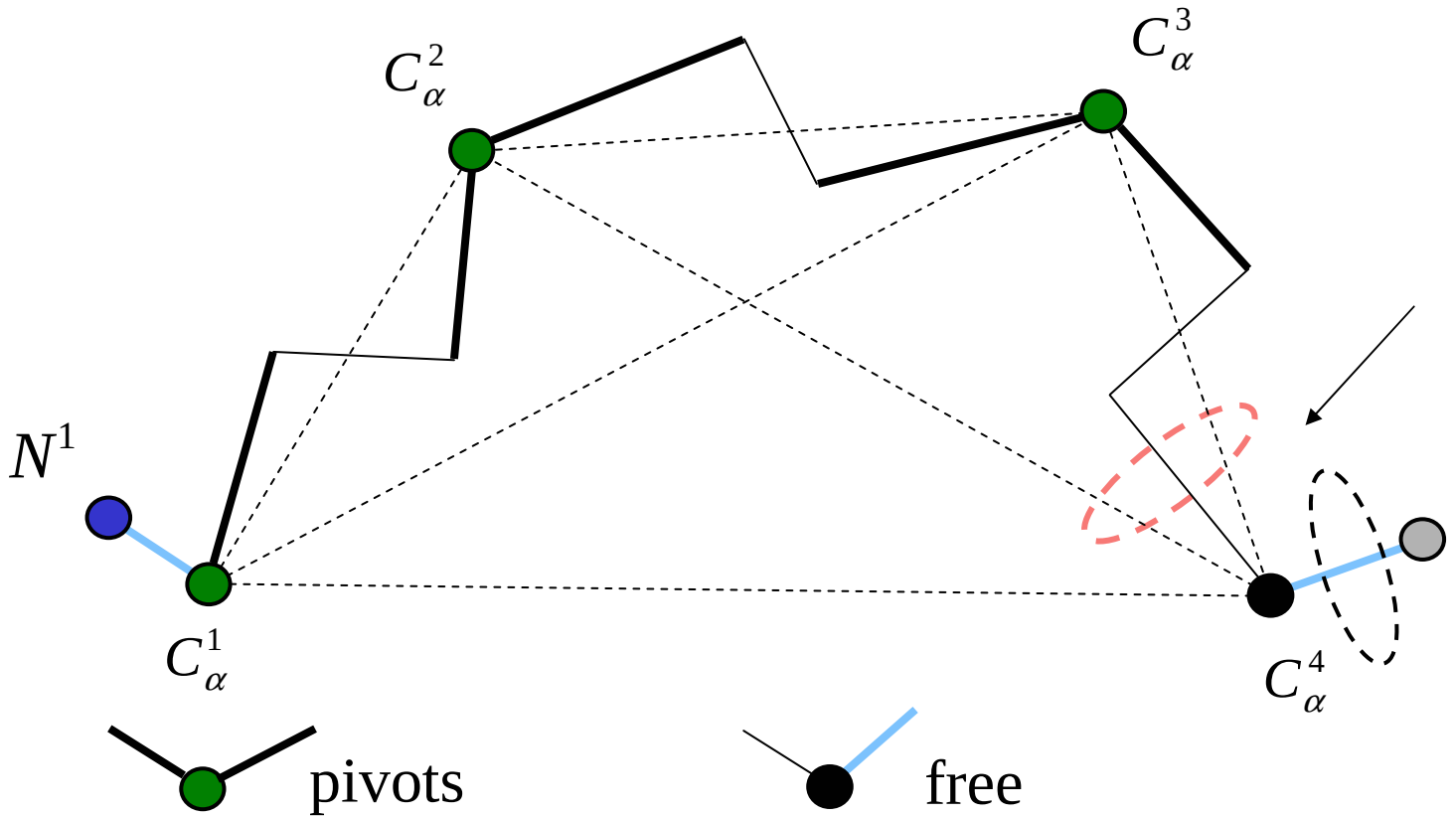


Special case: tailoring a cysteine bridge

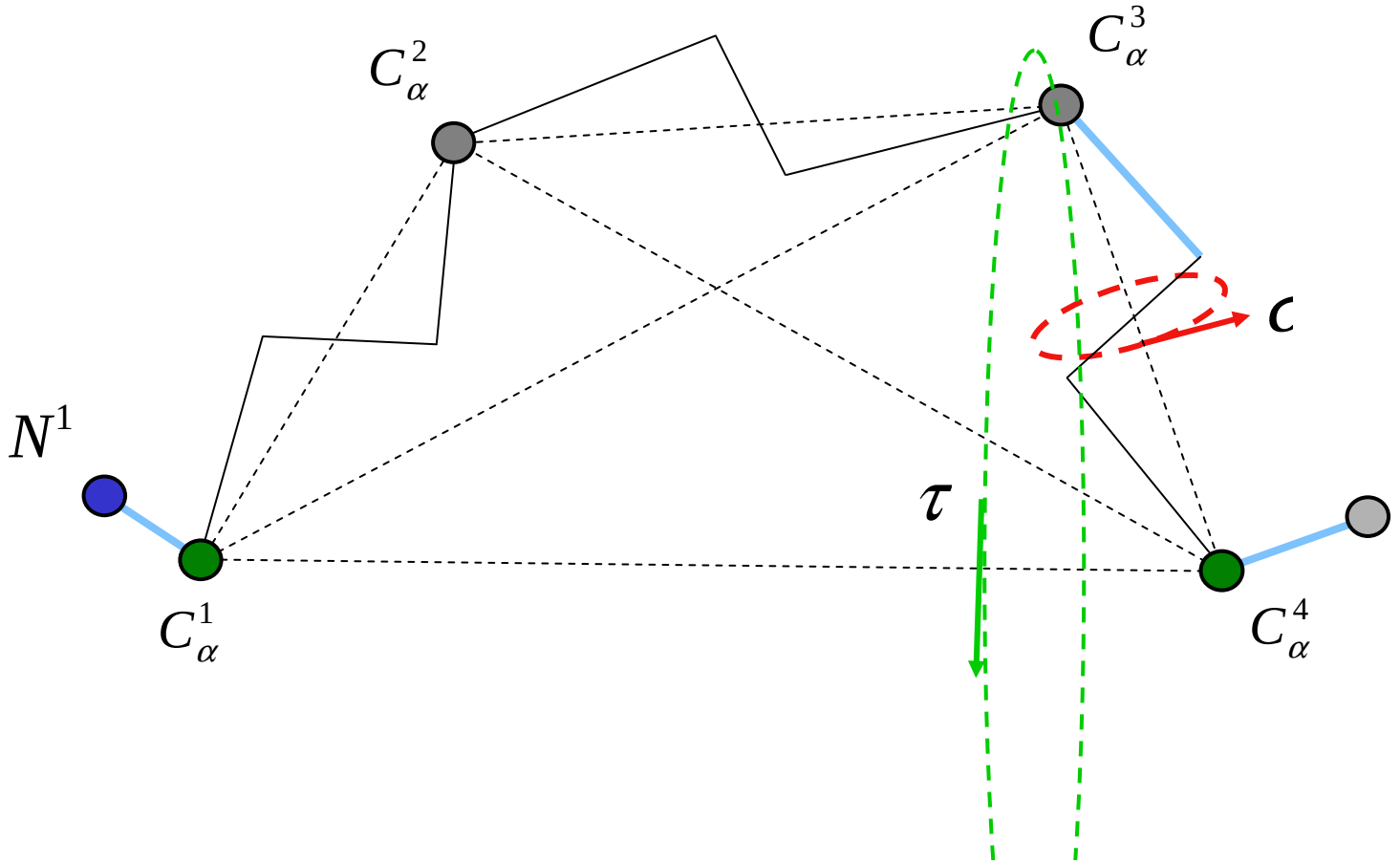

fixed in space
S2 bridge
free


fixed in space
virtual

(1a) Tetrapeptide: 2 DoF



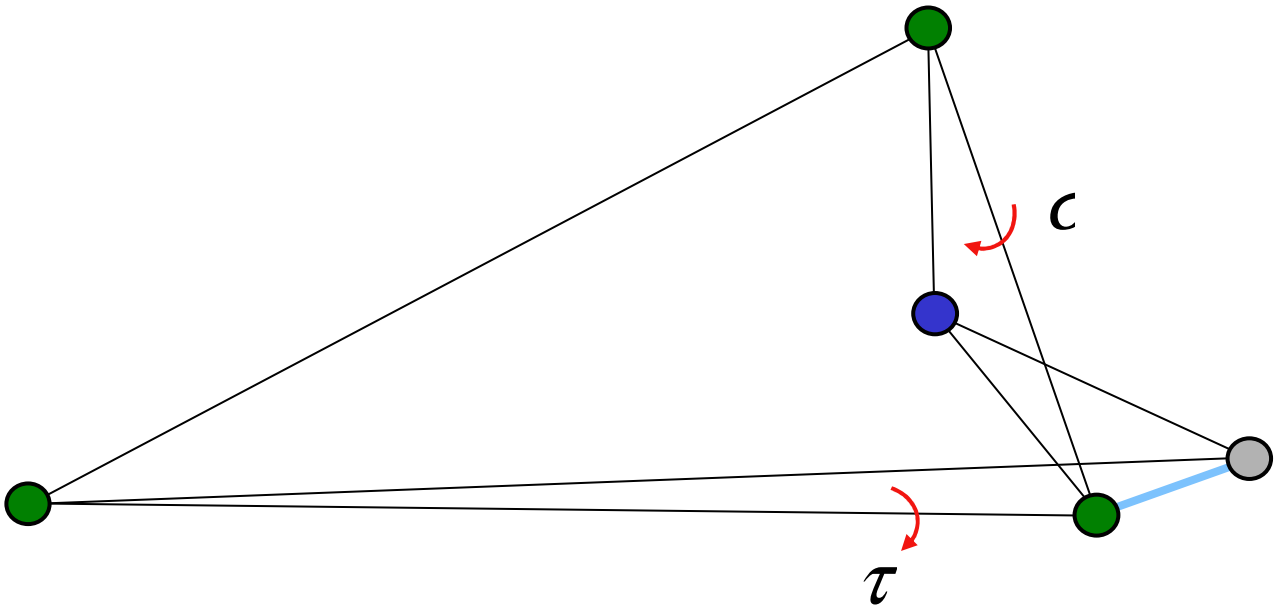
(1b) tetrapeptide with Ca1-Ca3 distance fixed: 1 dof



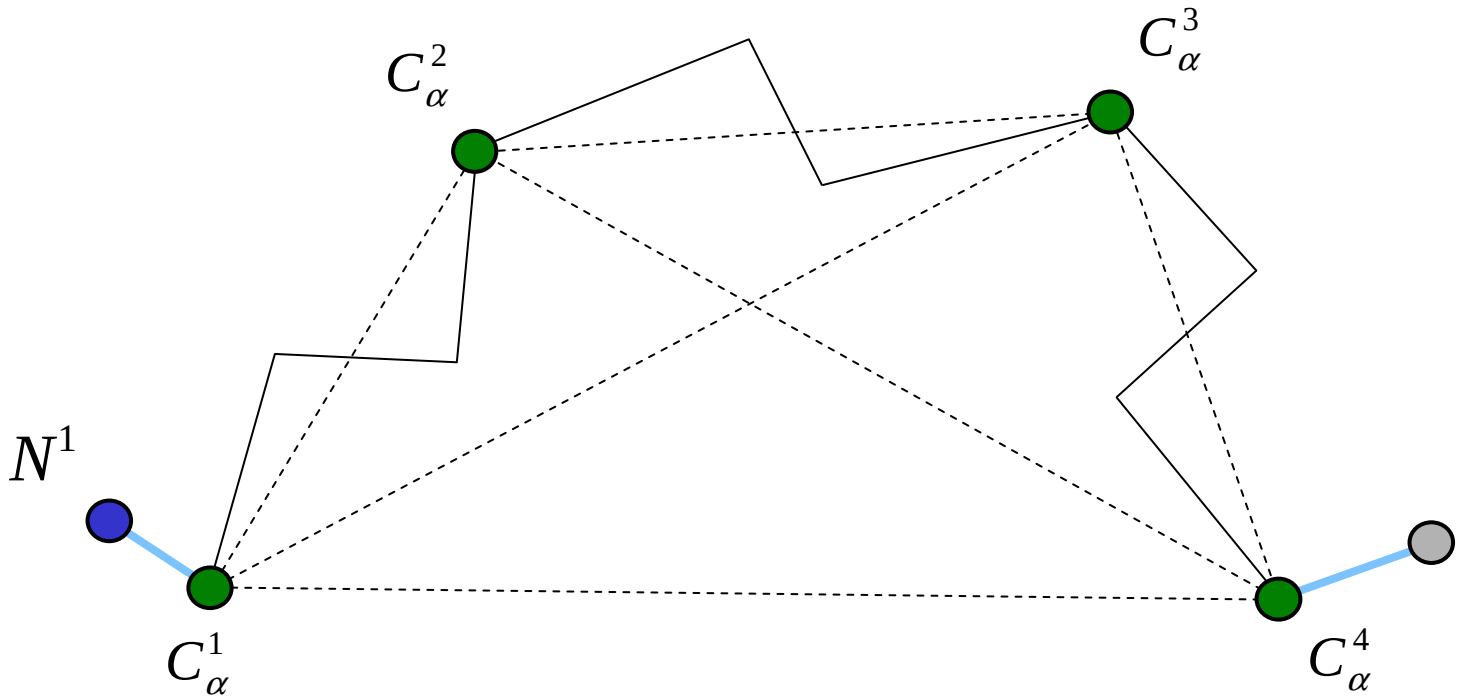
Tetrahedral equation

$$G(\sigma,\tau) := Au^2v^2 + Bu^2 + Cuv + Dv^2 + E = 0$$

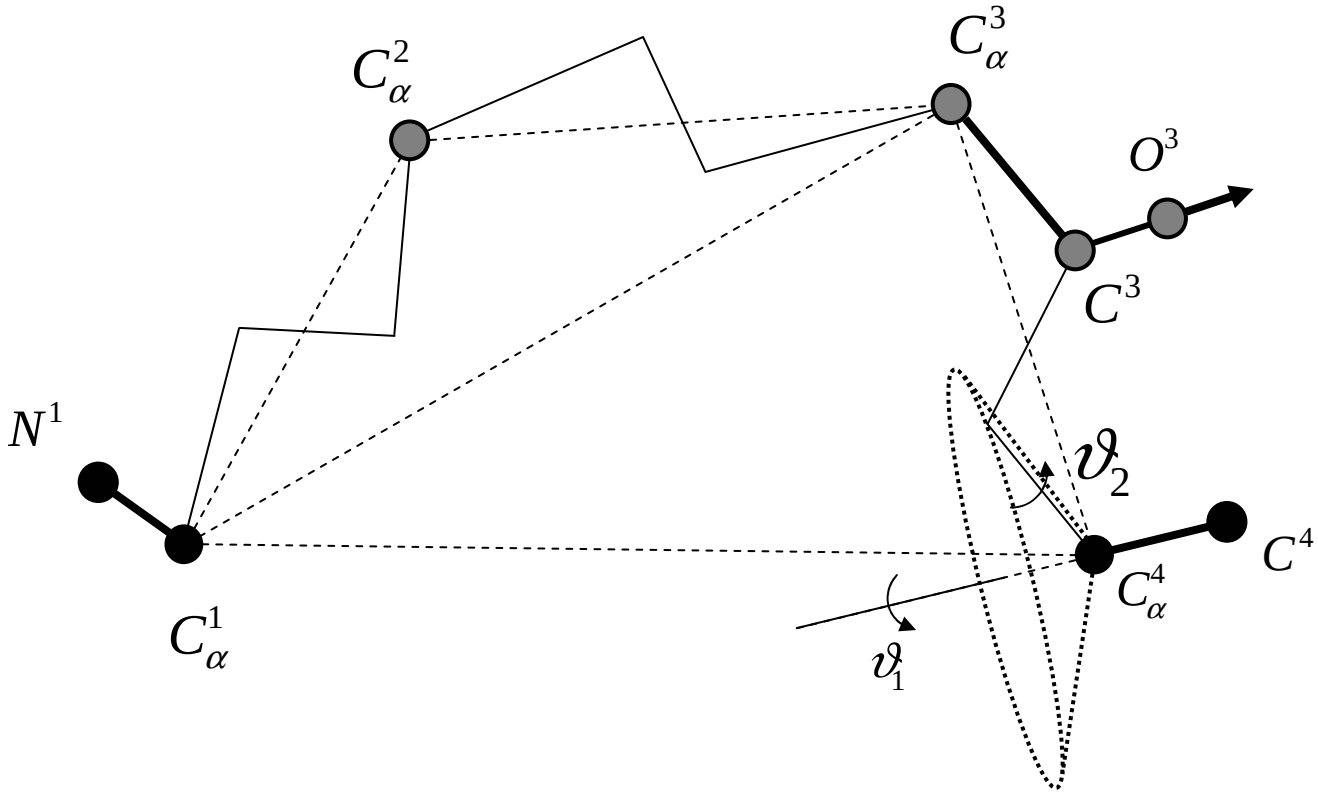
$$u = \tan(\sigma/2); v = \tan(\tau/2)$$



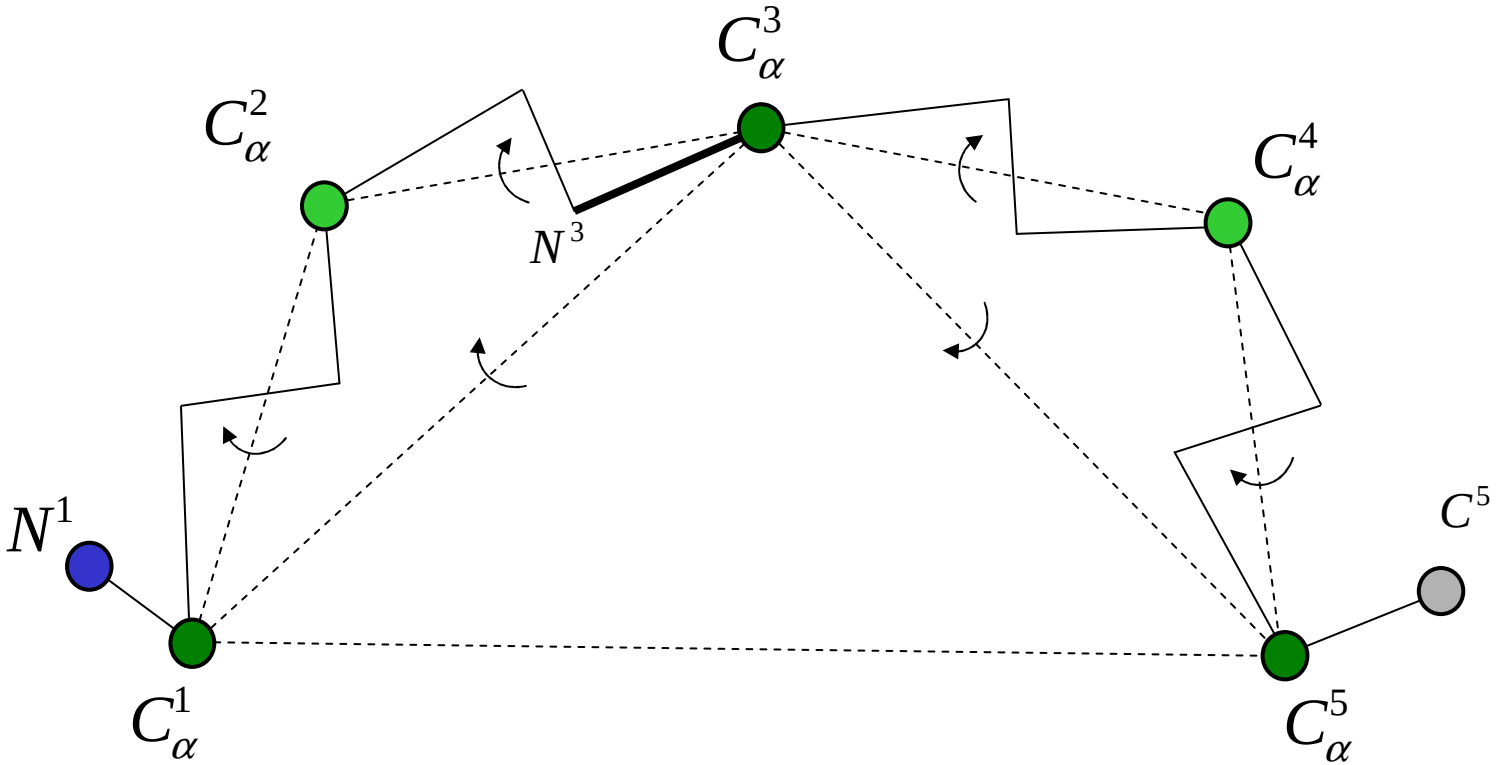
(1c) tetrapeptide with Ca1-Ca3 and Ca2-Ca4 distances fixed: up to 32 solutions (4 tetrahedral equs, 0 dof)



(2) A tetrapeptide with a directional constraint: here a CO bond is constrained to hydrogen-bond forming range

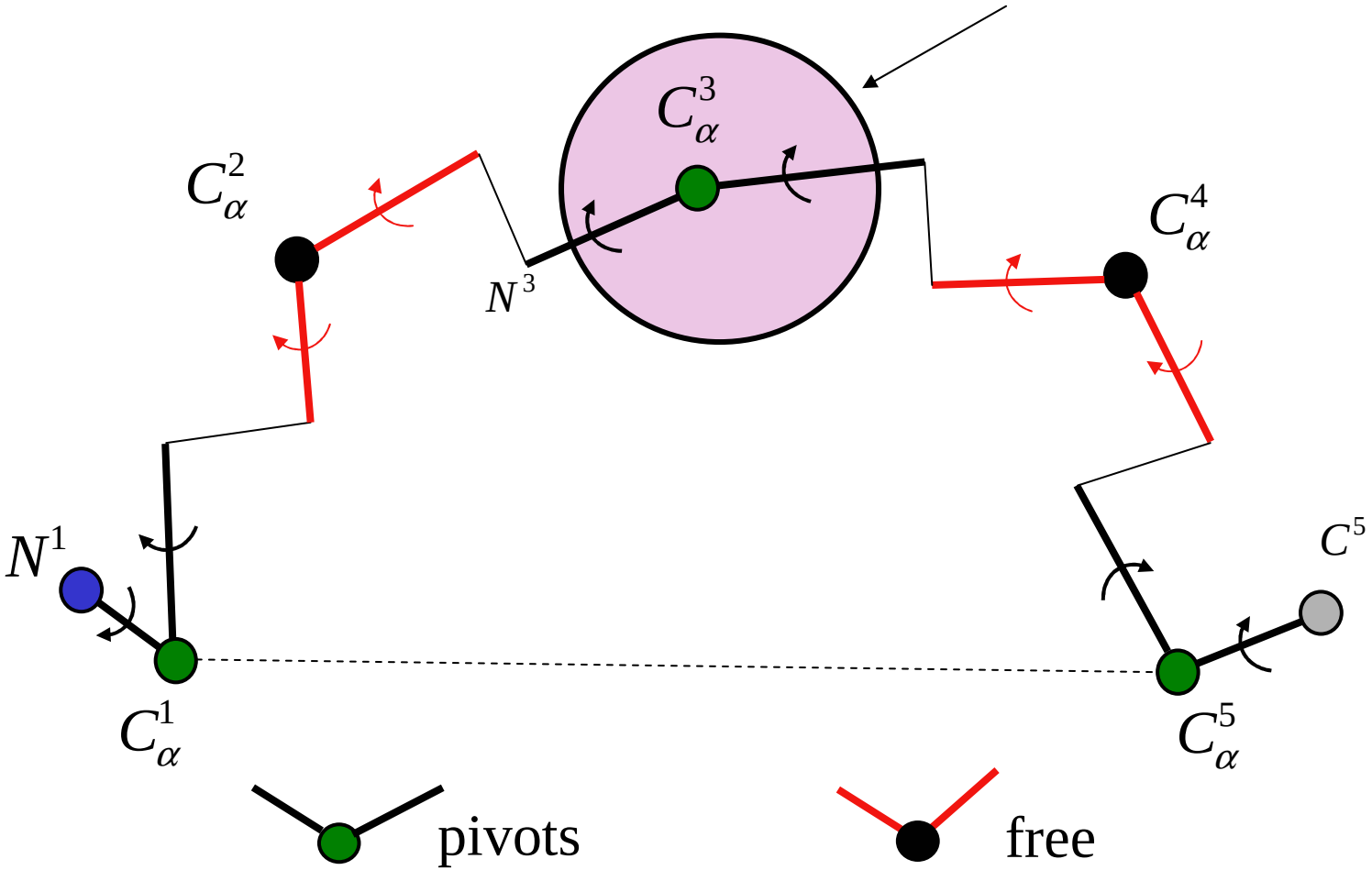


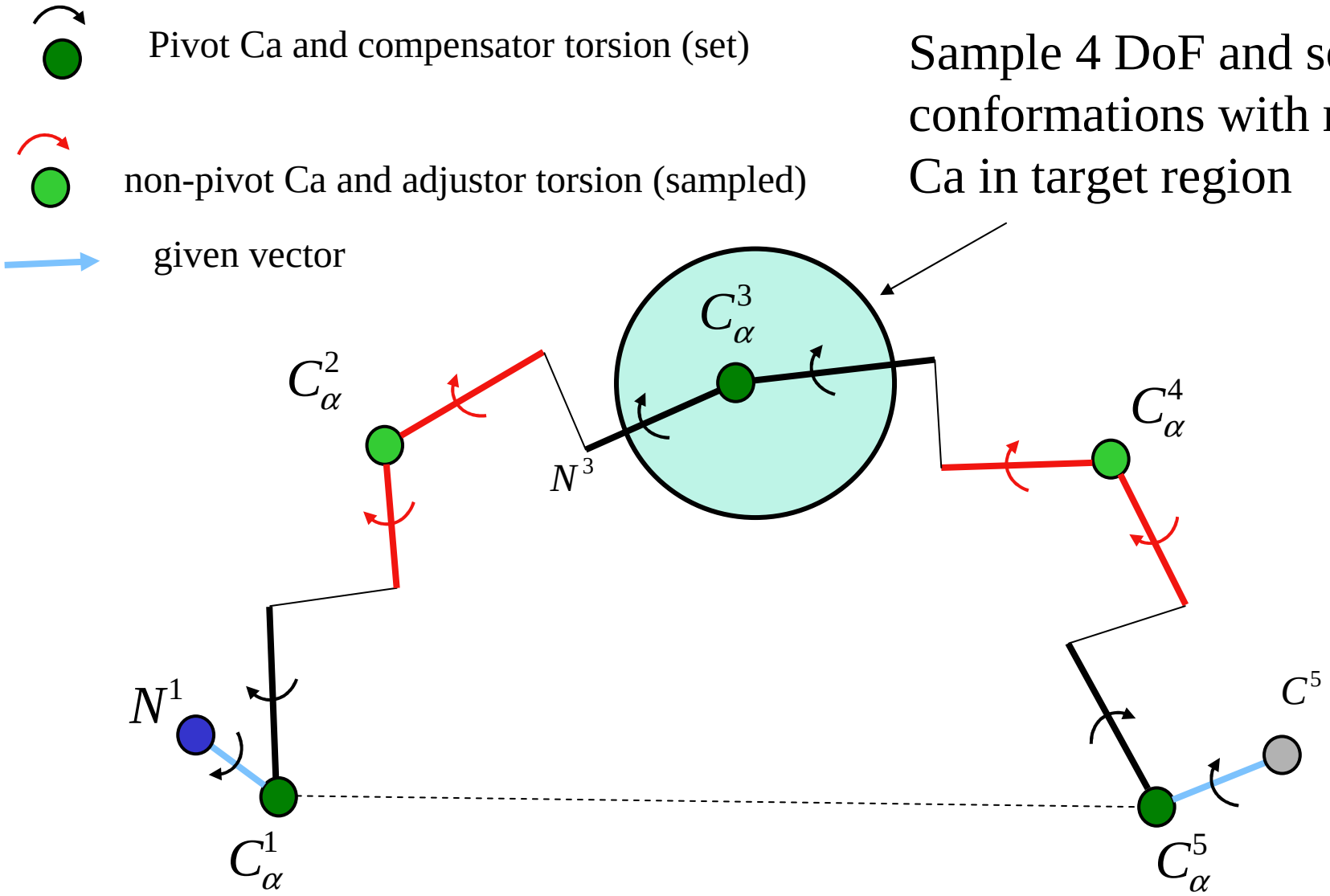
(3) Localization constraint: Pentapeptide loop with mid-Ca atom fixed



Constraint count: 2x5 – 6(closure) – 3(localization)= 1 DoF

(3a) Pure sampling: 4DoF sampled, filter for mid-Ca in target region

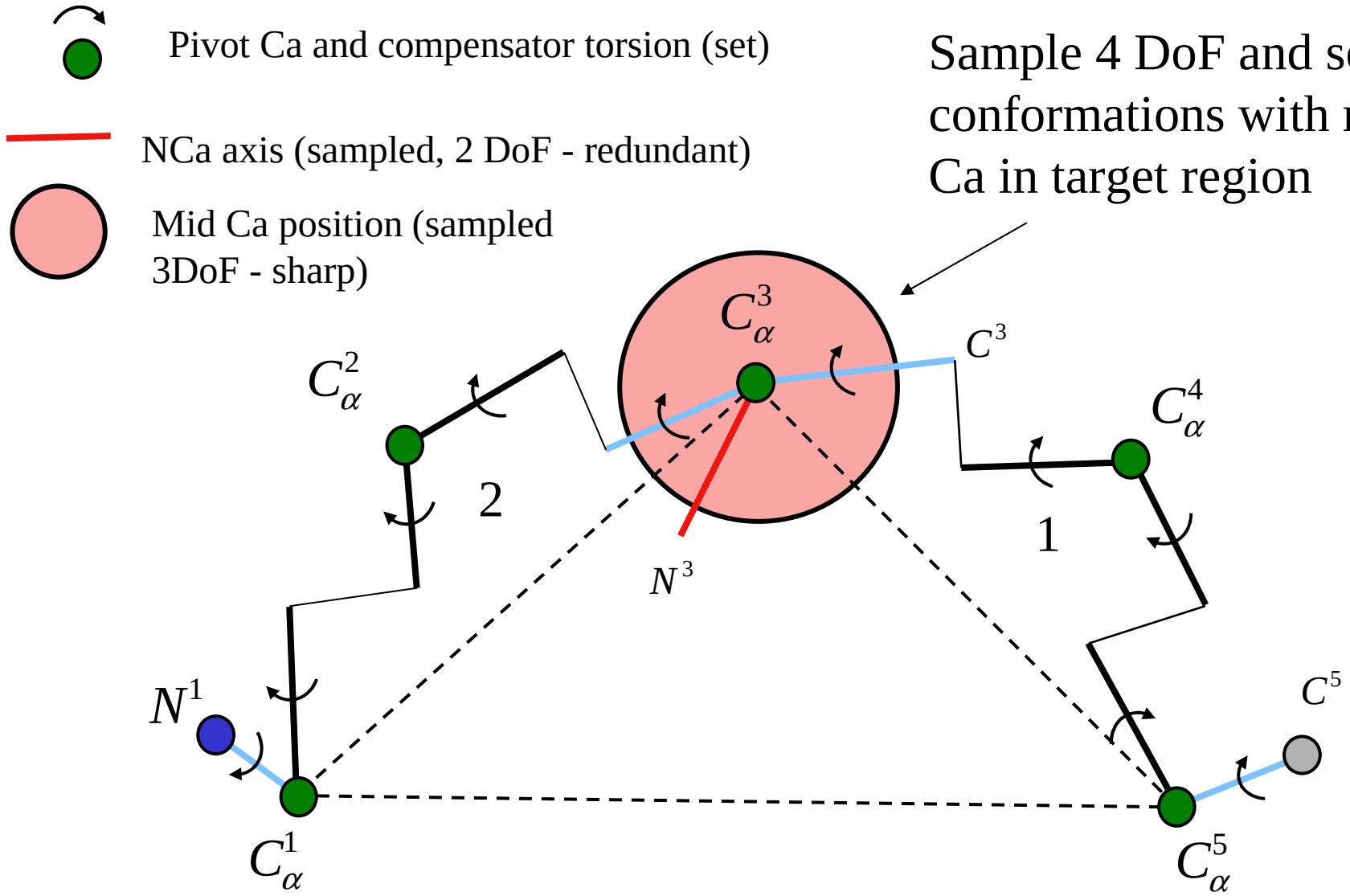




Pure sampling: 4DoF sampled, filter for mid-Ca in target region

(3b)Localization: double loop-closure

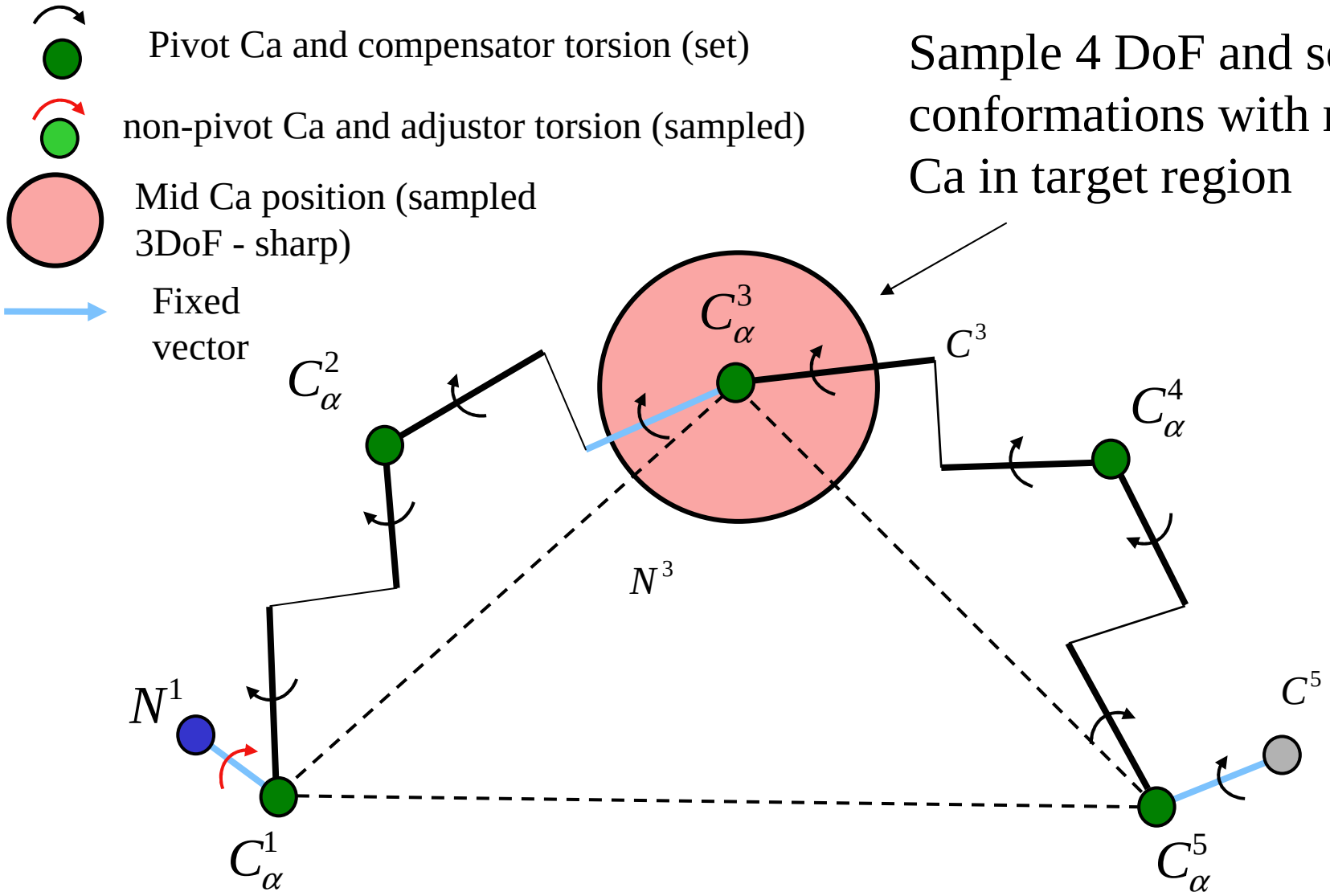
- First closure: set mid-Ca to a point in target region (restricted 3d search)
- Choose a NCa vector (full 2d search)
- First closure: based on res. 3-4-5, fixes CaC
- Second closure: based on res. 1-2-3, moves NCa to new position
- All atoms placed and Ca correct, but need to select based on feasibility of Cb. Cannot easily avoid redundant sampling.



LC+position (Ca: 3 cartesian coord) /orientation (NCa: 2 angles)

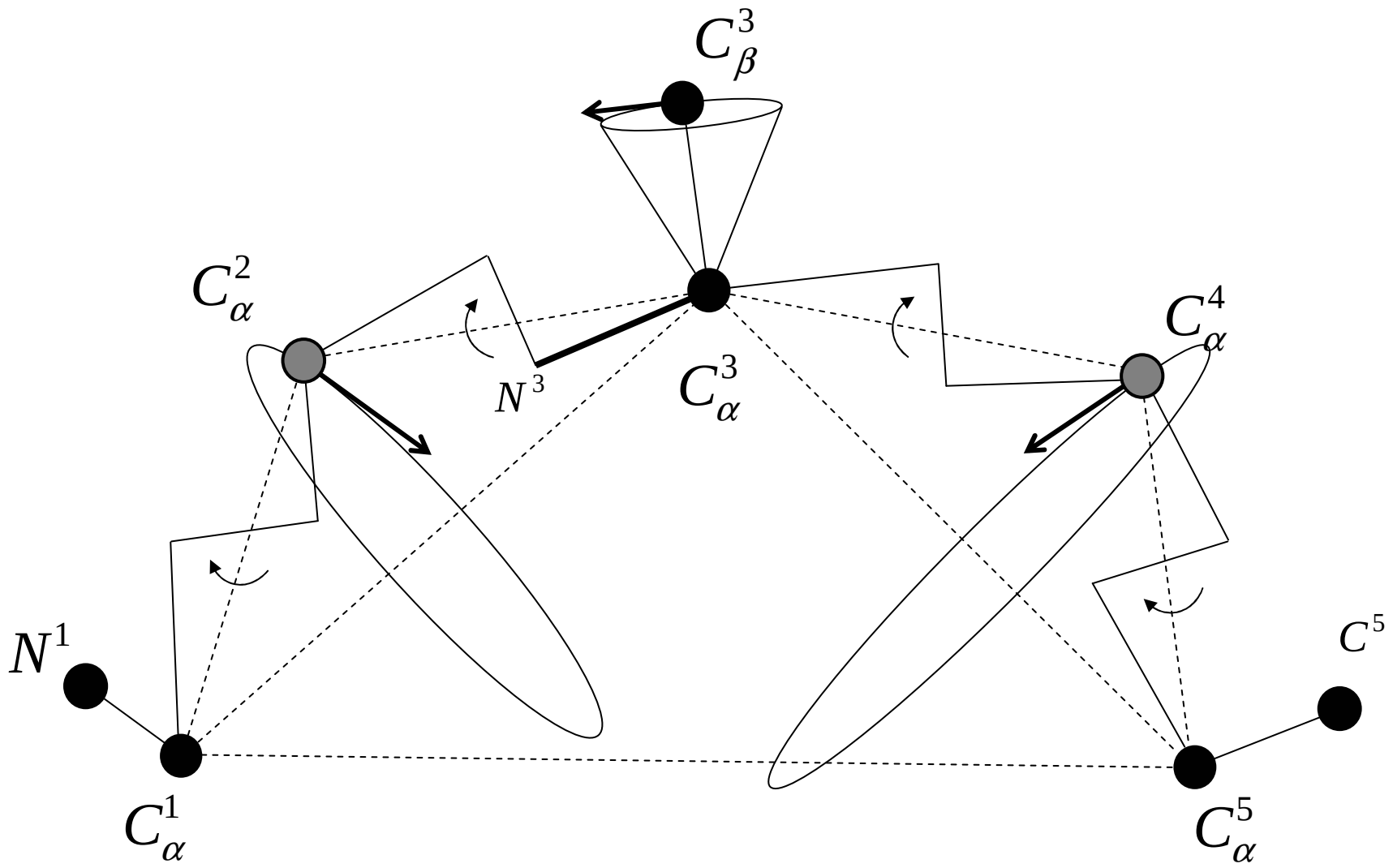
5DoF sampling: prescribed localization / oversampled orientation

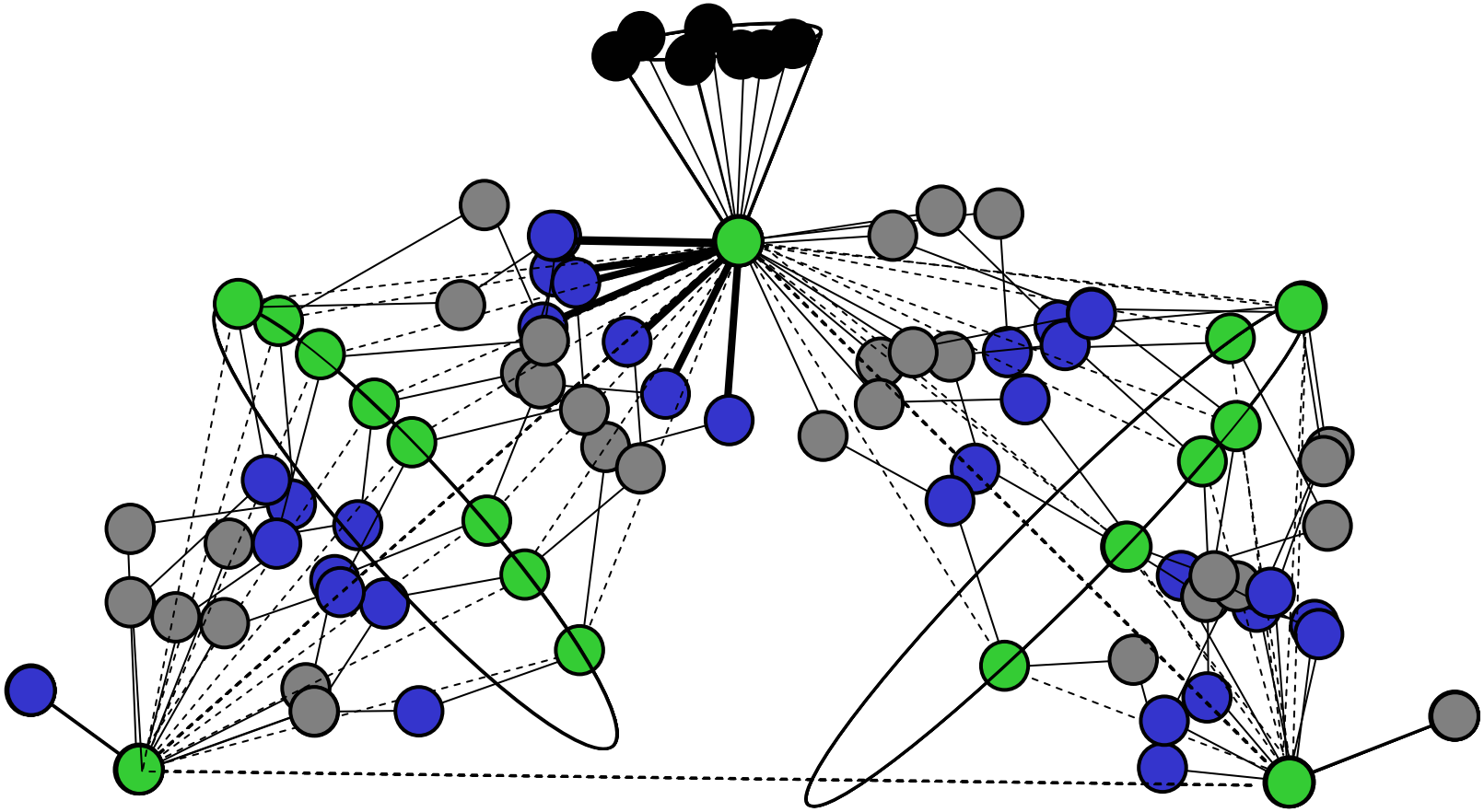
(3c) Optimal strategy (1dof sampled)

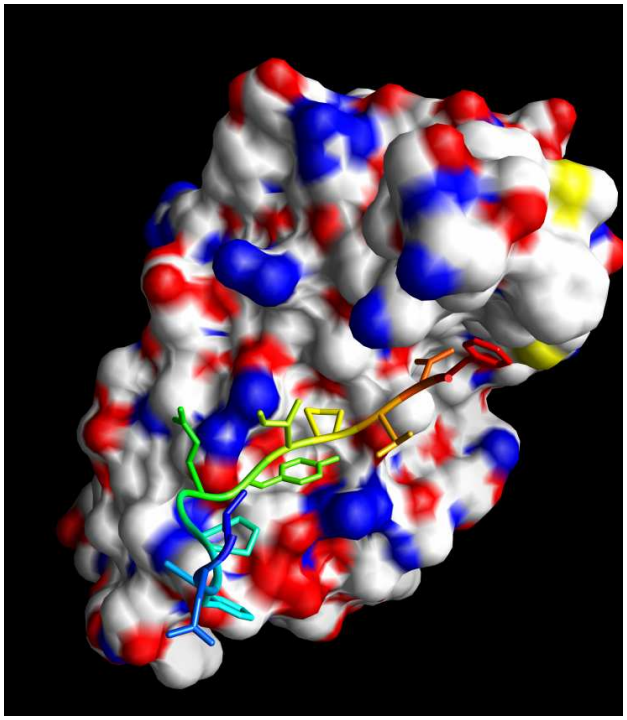


LC+position (Ca: 3 cartesian coords) /orientation (NCa: 2 angles)

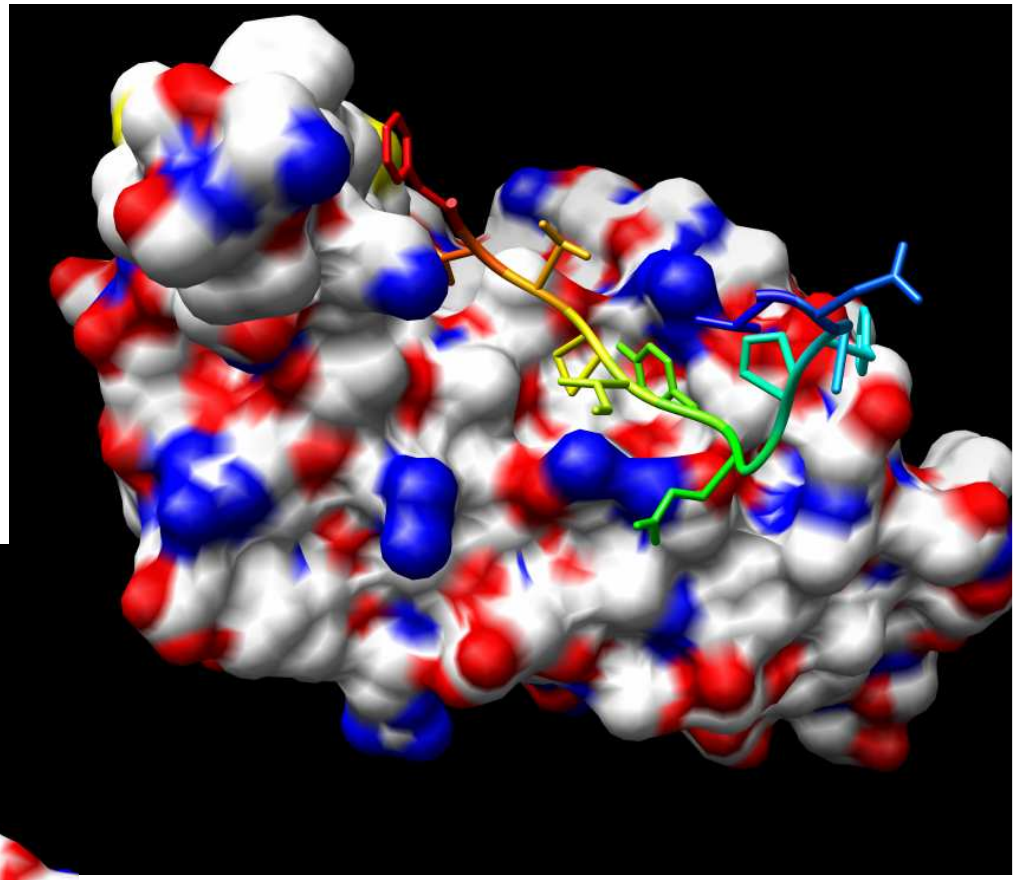
5DoF sampling: prescribed localization / oversampled orientation

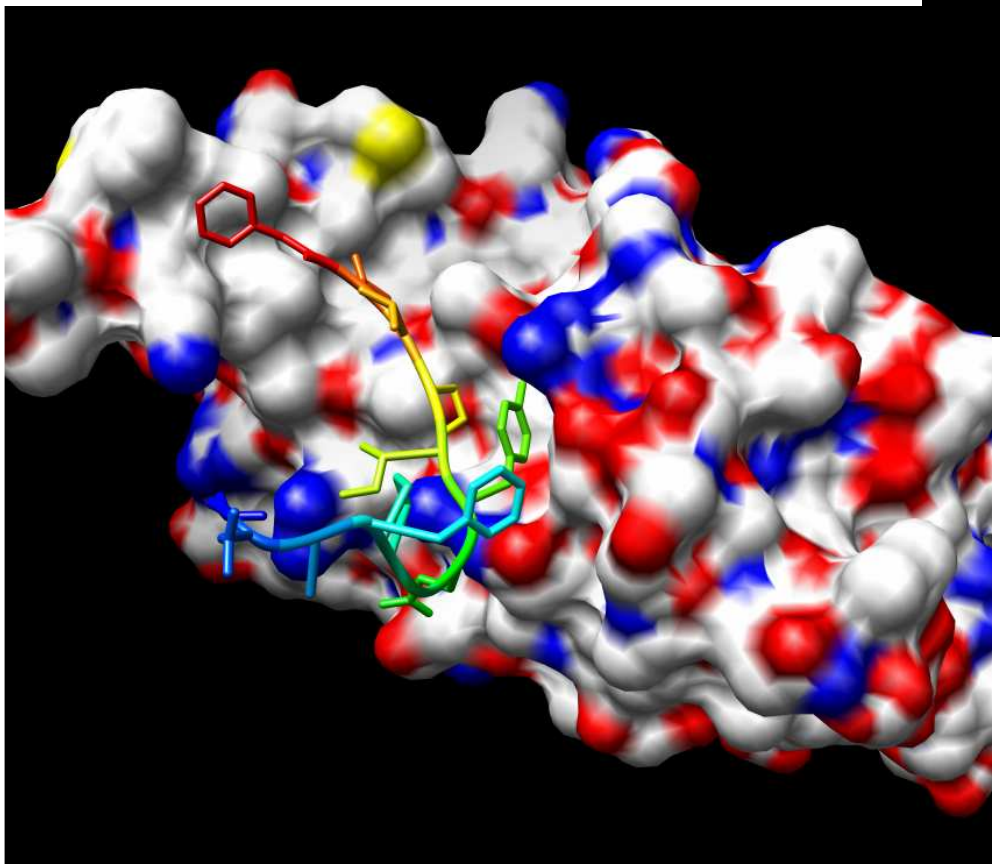

C_β^3
C_α^2
N^3
C_α^3
C_α^4
N^1
C^5
C_α^1
C_α^5



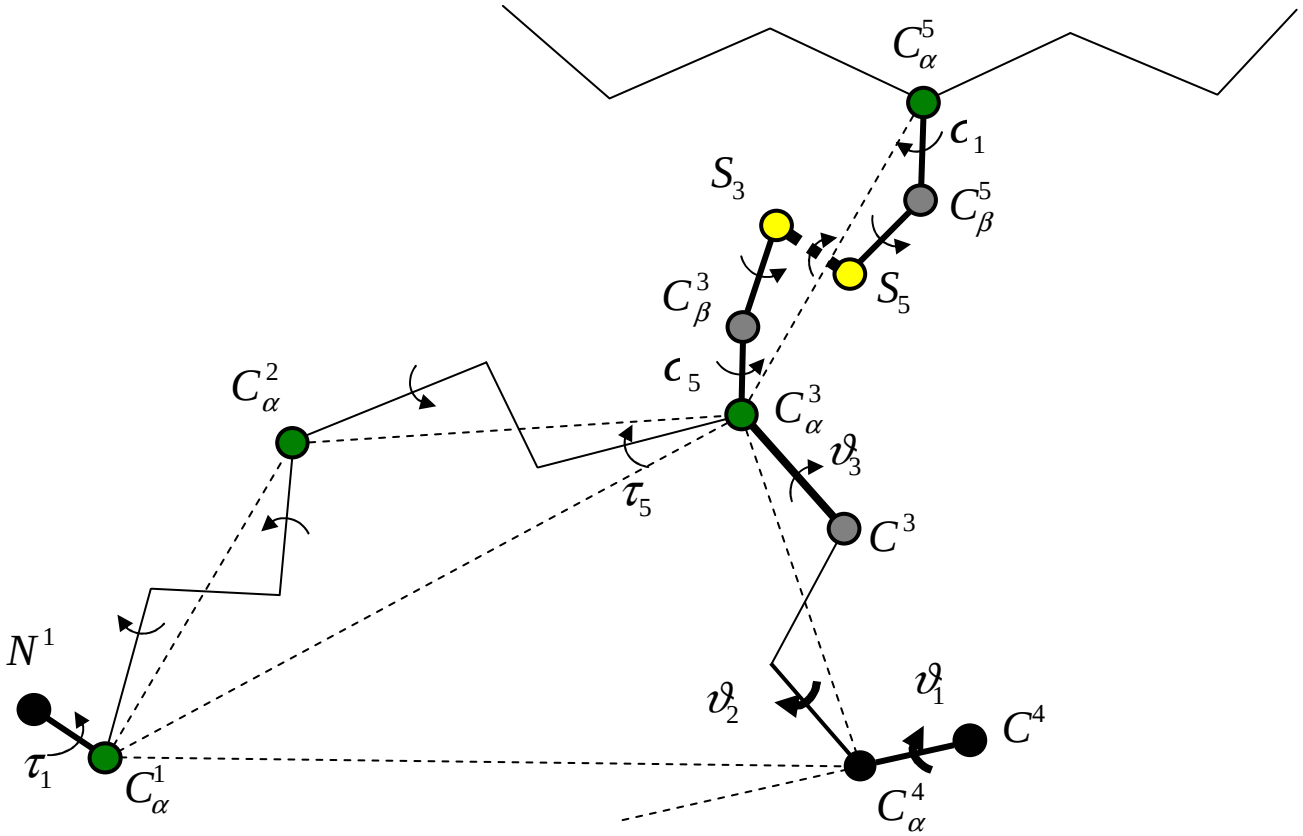


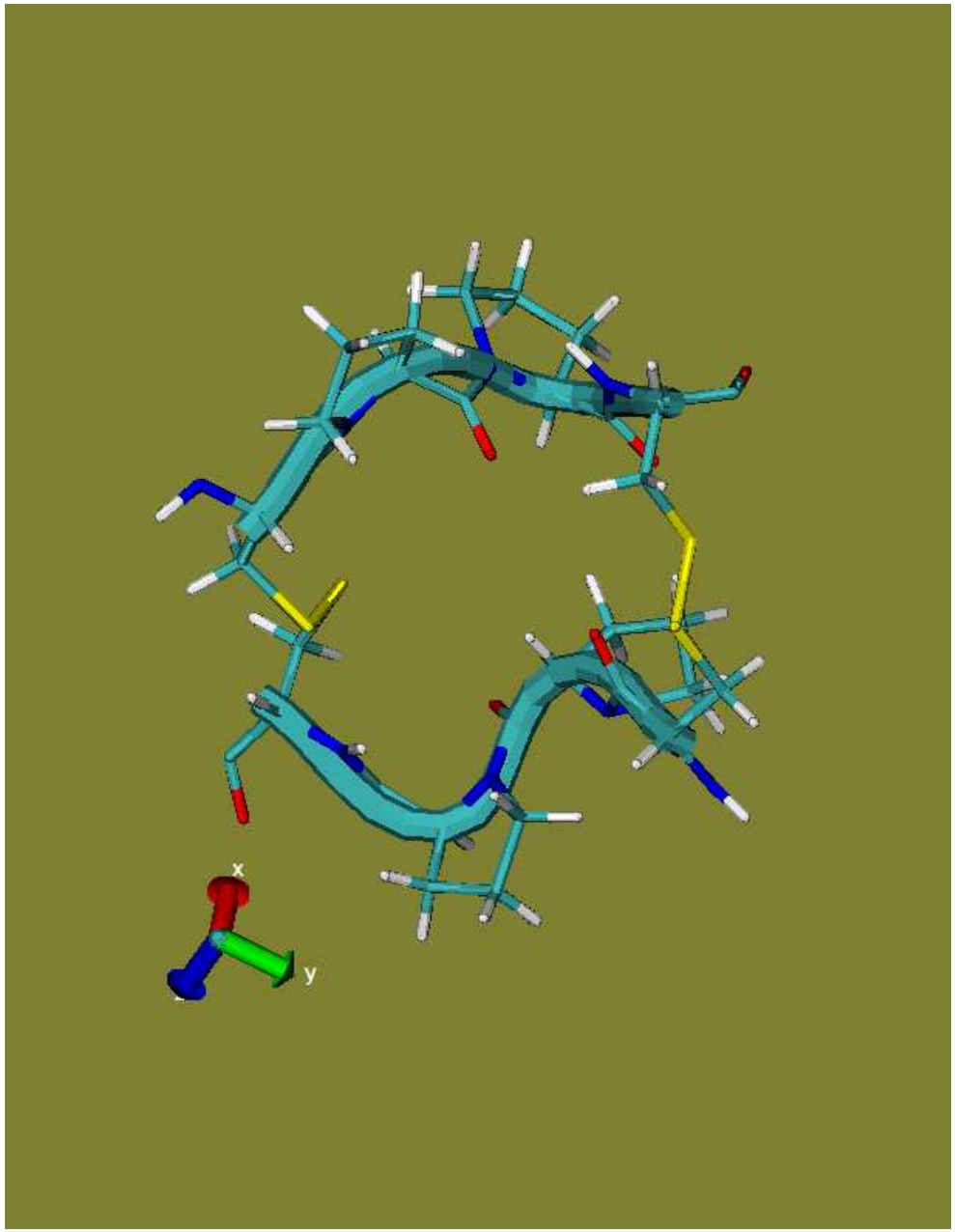
C_{β}^{3}
C_{α}^{2}
N^{3}
C_{α}^{3}
C_{α}^{4}
N^{1}
C^{5}
C_{α}^{1}
C_{α}^{5}



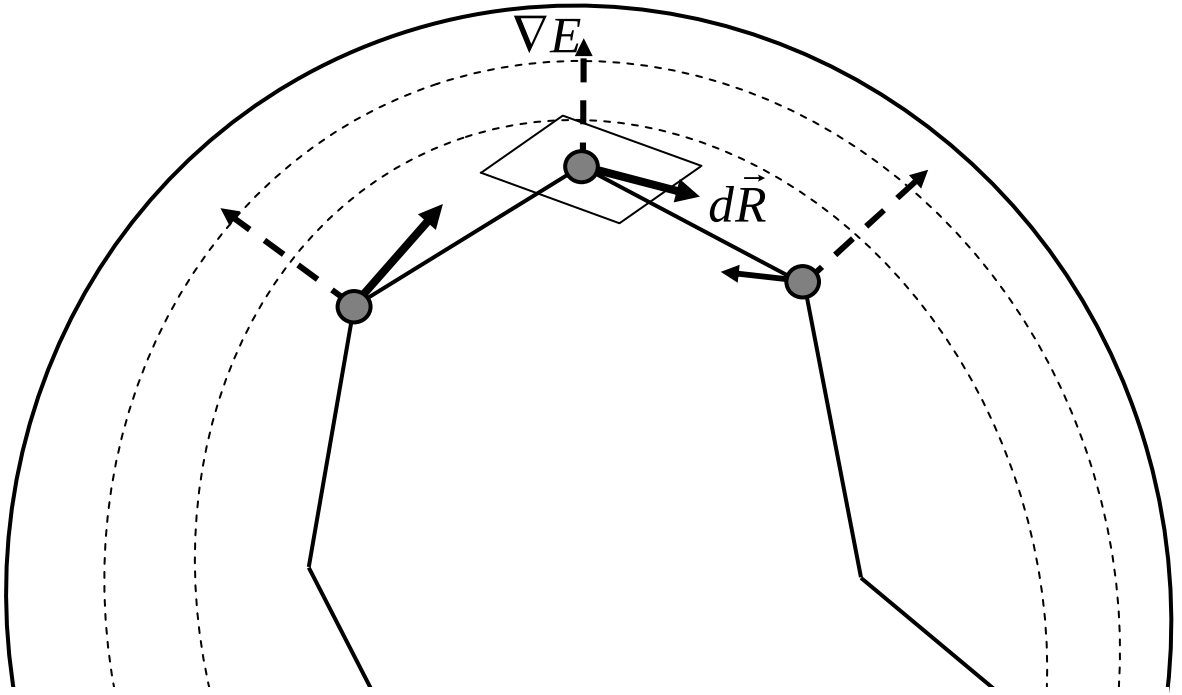


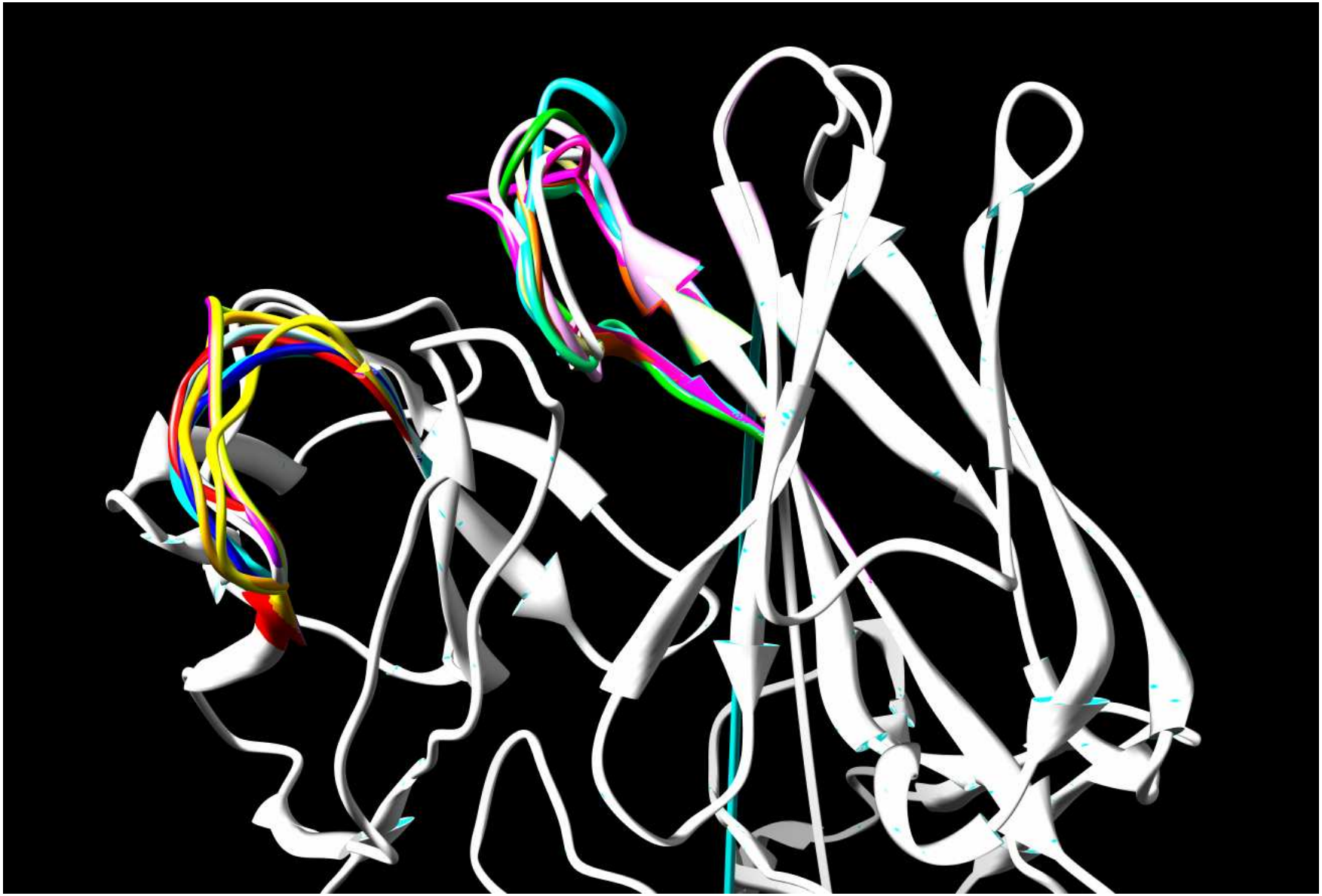
(4) Tetrapeptide loop with a Cysteine bridge to fixed backbone.




x
y

(5) Guided sampling: Exploring conformations compatible with a binding pocket.





Acknowledgements

- Ken Dill, Matthew Jacobson, Tanja Kortemme, Dan Mandell, UCSF
- Michael Wester, Dept of Math and High Performance Computing Center, UNM
- Shawn Martin, Mike Brown, Jean-Paul Watson, Sandia Labs
- Bob Lewis, Fordham U., Manfred Minimair, Setton Hall U.
- Sara Pollock, Dept of Math and Division of Biocomputing, UNM (now at UCSD)
- Chaok Seok, Dept. of Chemistry, Seoul National University, S. Korea
- Support: NIH

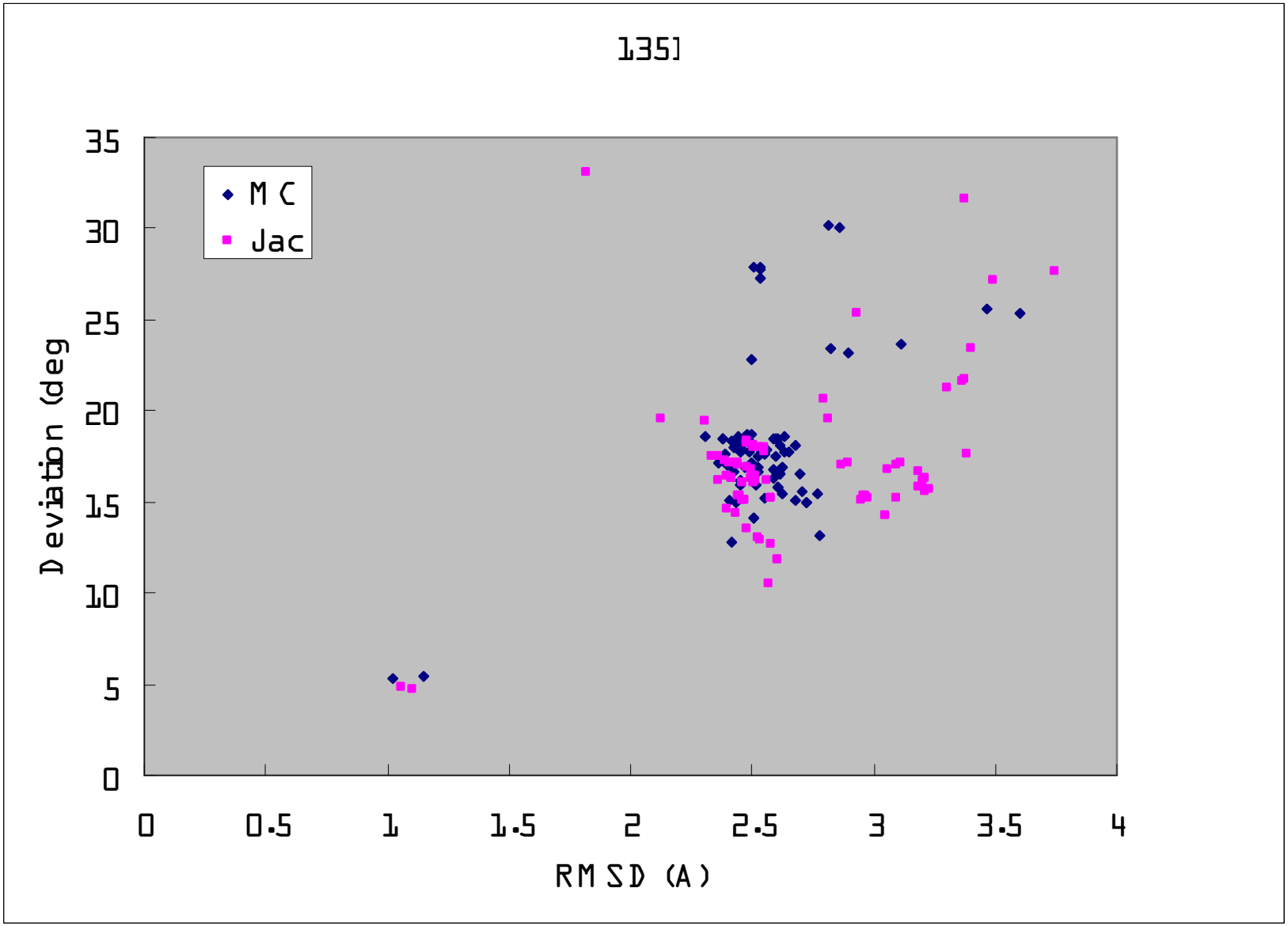
Appendix: Other applications

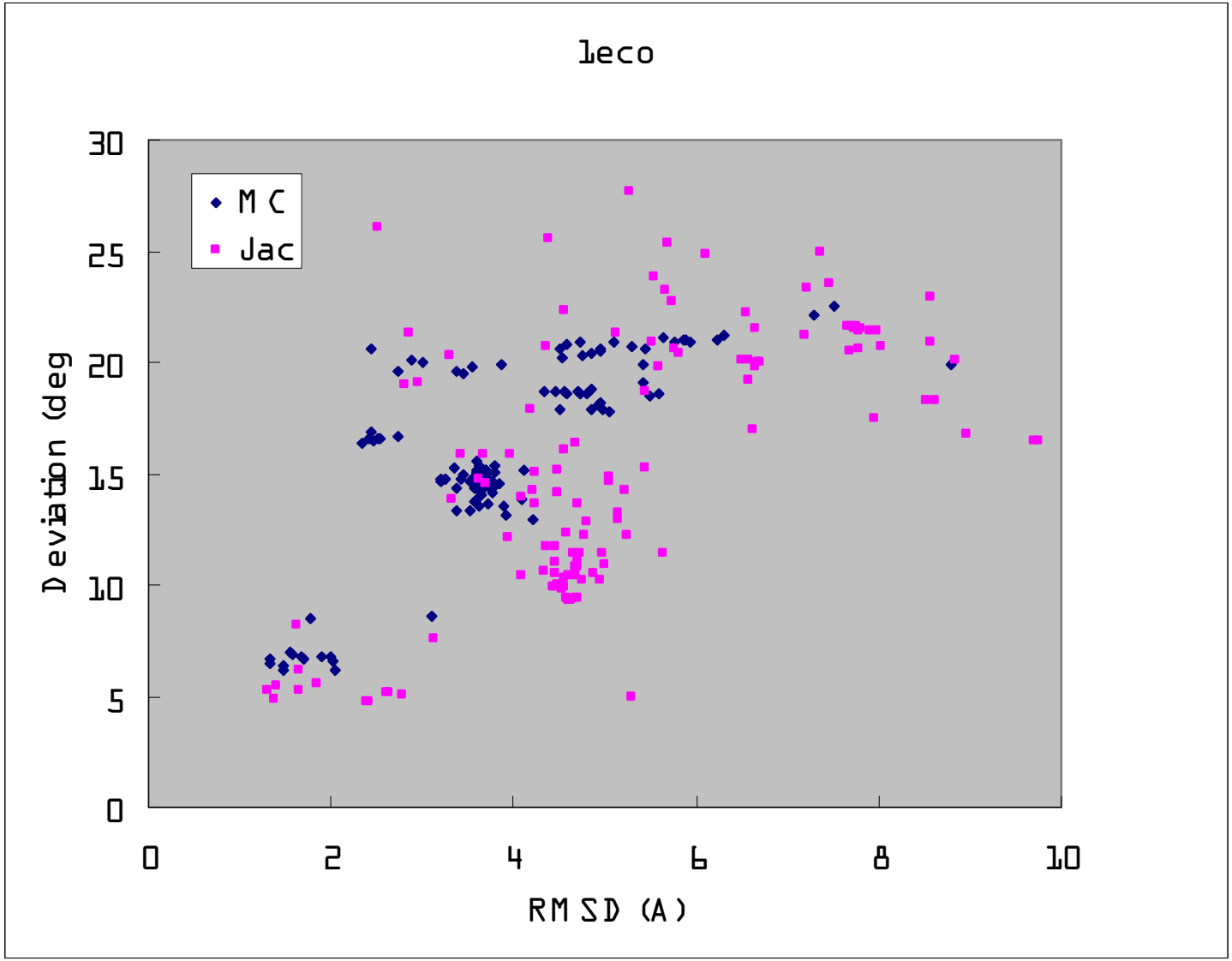
- Fragment assembly (Chaok Seok, Joulyan Lee)
- Concerted moves Monte Carlo (Jerome Nilmeier, Matt Jacobson, Lan Hua)
- Helical protein assembly (Albert Wu, Ken Dill, Justin MacCallum)

Minimization of Angle Deviation

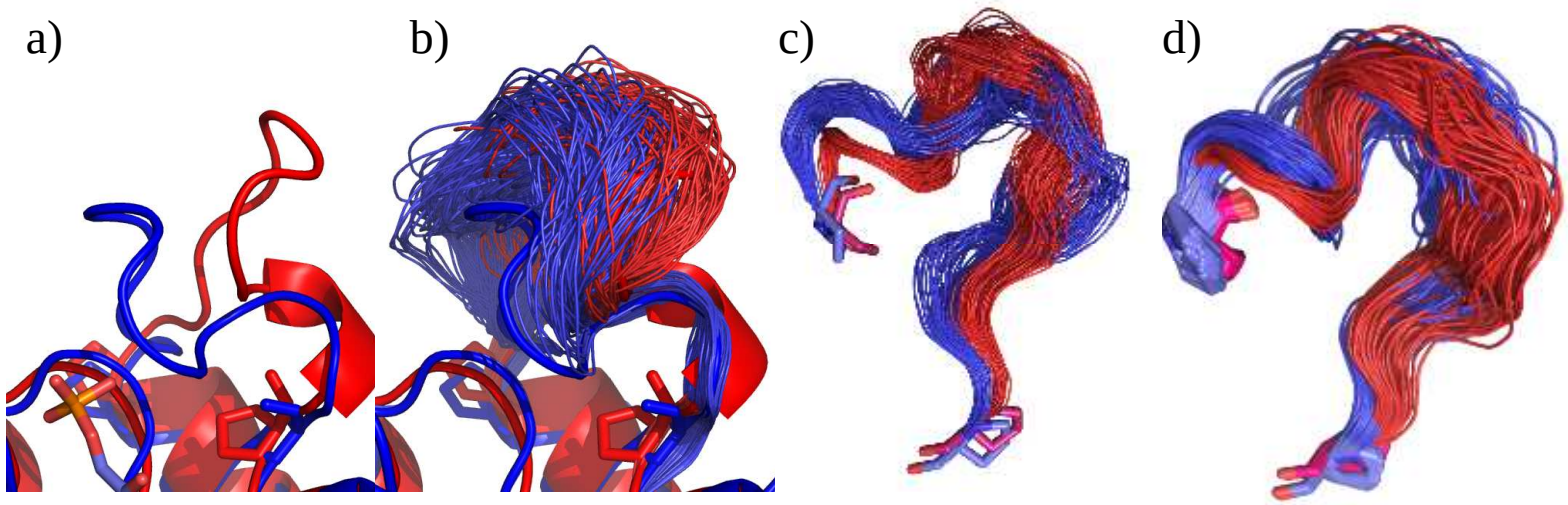
- Fragment assembly methods are often applied to protein structure problems.
- When structures are generated for a segment such as a protein loop, assembled fragments are not fit into the frame of the protein of interest exactly.
- Deviation of the dihedral angles of the loop from the fragment angles are minimized here to maintain the features obtained from the structure database as well as possible,
- Two methods have been tried: 1) Monte Carlo simulation and 2) A local minimization in the space of loop conformations satisfying the loop closure constraint. In both method, root-mean-square deviation in dihedral angles is used as the objective function.
- Monte Carlo simulation: 1 driver angle is perturbed randomly within 10 deg, 6 torsion angles are used to close the loop. kT=0.5 deg and 2000 MC steps. 20 independent simulations starting from different initial loop closure were performed for each starting conformation generated from fragment assembly.
- Minimization using the kinematic Jacobian: 100 steps of steepest descent minimization, and subsequent LBFGS-b minimization (termination criterion: function decrease: 10^7*machine precision, gradient: 10^(-3)). 20 independent minimization starting from different initial loop closure.
- Results: two loops, 8-residue loop of 135l (residues 84-91) and 12-residue loop of (35-46), were tested. R_ave is RMSD averaged over the different conformations generated from fragment assembly. R_min is the min RMSD. Dphi_init is the initial deviation in angles, and dphi_ave is the final deviation averaged over the conformations. The overall performance of the two methods is similar, but the computation time is much faster with Jac method.

Protein		1351	1eco
length		8	12
# of conf from fragment assembly		87	127
dphi_init (deg)		34.3	28.0
R_ave (A)	MC	2.5	3.9
	Jac	2.7	5.2
R_min (A)	MC	1.0	1.3
	Jac	1.1	1.3
dphi_ave (deg)	MC	18.1	15.8
	Jac	17.0	15.4
Time (sec)	MC	38.9	91.7
	Jac	5.4	5.1


135]
35
30
25
20
15
10
5
0
Deviation (deg
M C
Jac
0
0.5
1
1.5
2
2.5
3
3.5
4
RMSD (A)

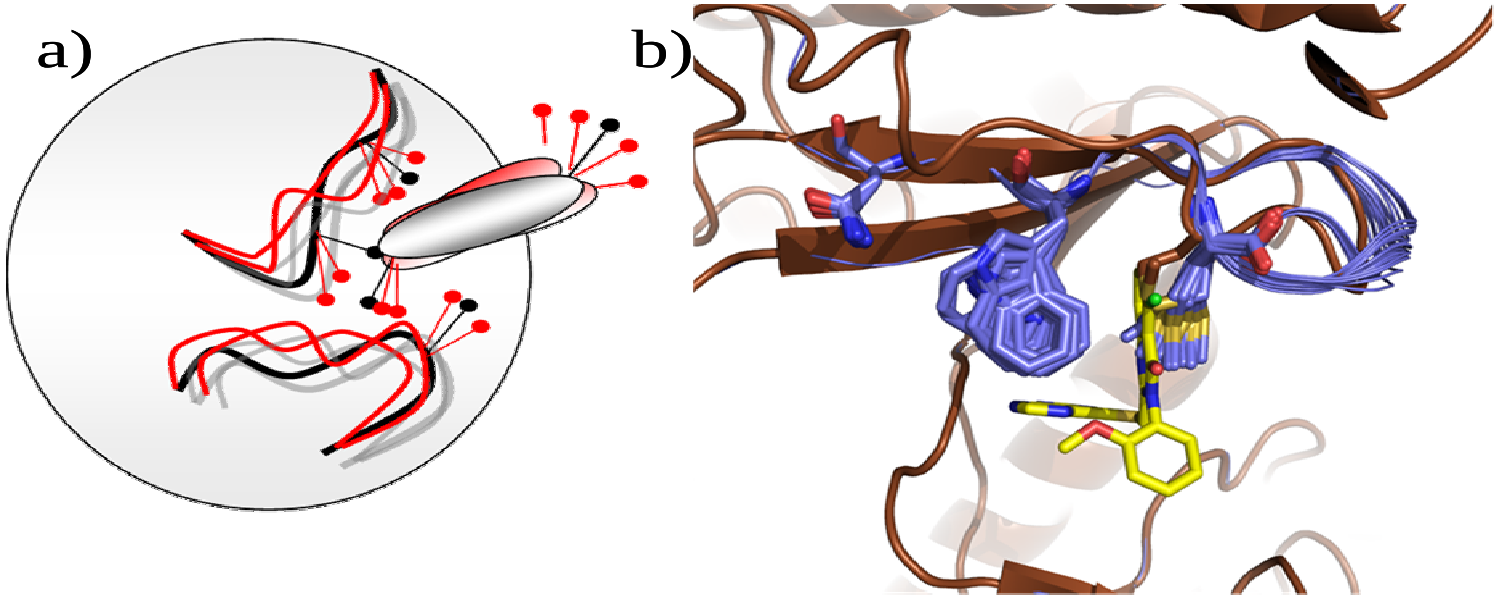

1eco
MC
Jac
Deviation (deg
RMSD (A)
0
5
10
15
20
25
30
0
2
4
6
8
10

TIM loop Dynamics using TLC



a) Native closed loop (blue) bound to ligand PGA), and Native open loop (red).
b) apo simulations of loops.
c) Loop simulation without the use of proline loop closure (pucker) moves, and
d) Loop simulation with the incorporation of proline pucker trial moves

Constrained binding pocket simulations using loop closure



- a) schematic of a binding pocket as a series of loops.
- b) Preliminary simulation of PI3 kinase, with a single loop and adjacent sidechains.

Assembly of helical proteins

A simple heuristic based on fast loop closure and maximal hydrophobic packing—as measured by radius of gyration of the Ca atoms in hydrophobic residues

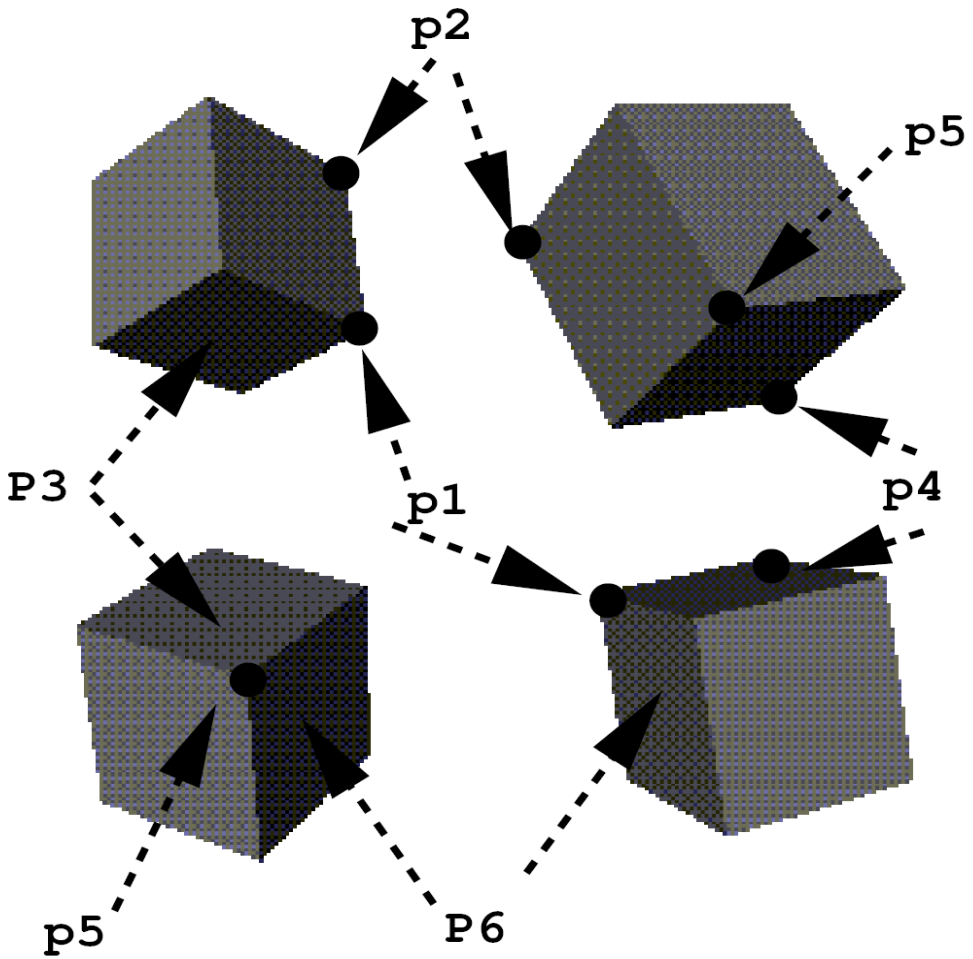
Motivated from need to improve assembly performance of Dill group's Zipping & Assembly strategy for tertiary structure prediction

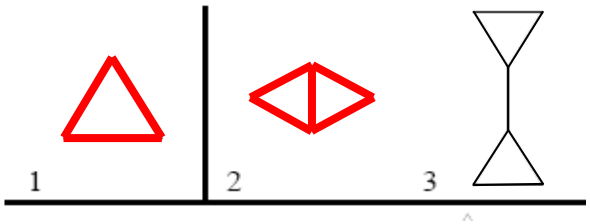
G.A. Wu, E.A. Coutsias, K.A. Dill, Iterative Assembly of Helical Proteins by Optimal Hydrophobic Packing, (Structure, 2002)

The assembly algorithm

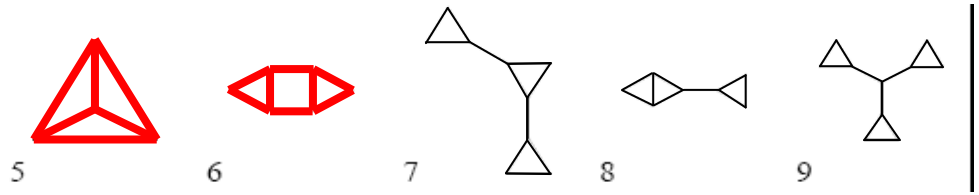
- Begin with a protein whose secondary structure is known to contain helices (as determined, e.g. by **DSSP**, Kabsch,Sander, Biopolymers 22, 2577-2637 (1983)) remove loops and consider the problem of placing the helices relative to each other
- Align two helices; score each alignment; select best subset, close loop(s).
- Align next helix with assemblage of first two, close loop; iterate
- Cluster/rank by RgH; select best candidates based on a hydrophobic packing criterion.

Object assembly en masse: identical equations to loop closure. A possible approach to avoid searching. Assemble all elements at once into geometrically feasible configurations.

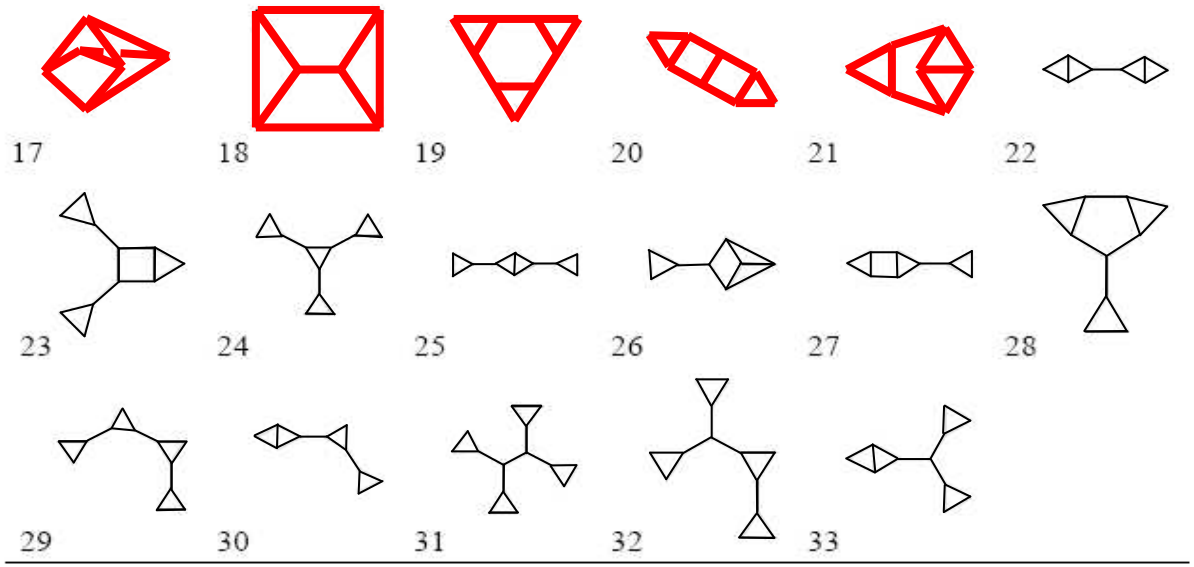




1, 2 rings



3 rings



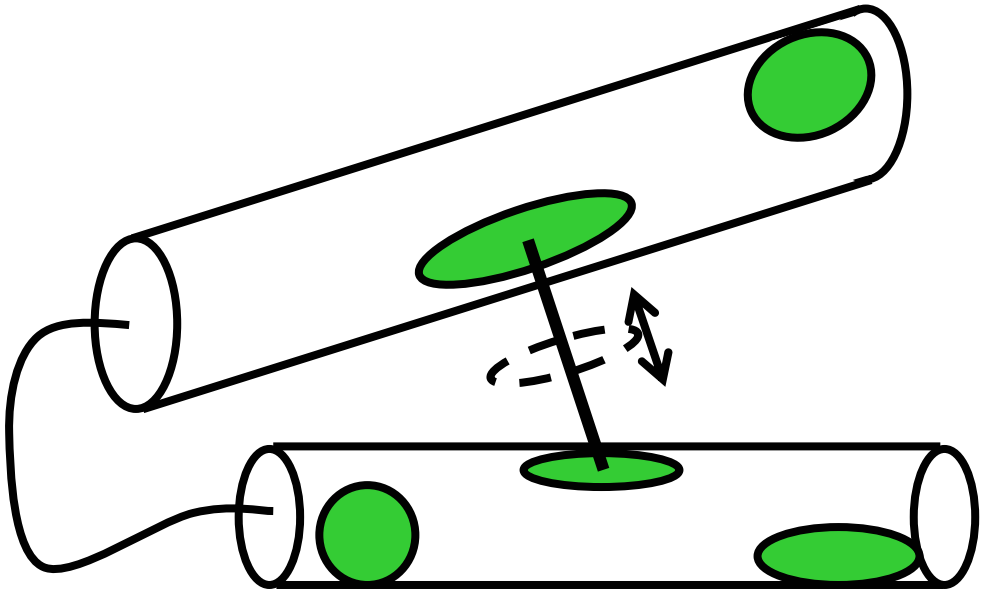
4 rings

all possible graph topologies with up to 4 rings, with only 3-node contacts

Wester, Pollock, Coutsias, Allu, Muresan & Oprea, Topological Analysis of Molecular Scaffolds II: Analysis of Chemical Databases (JSim, 2008)

http://topology.health.unm.edu

Graph Theory: assemble using topological graphs and combinatorics



Sample all possible hydrophobic pairings between two helices

2 DoF sampled: translation and rotation about alignment axis

Limited by loop closability

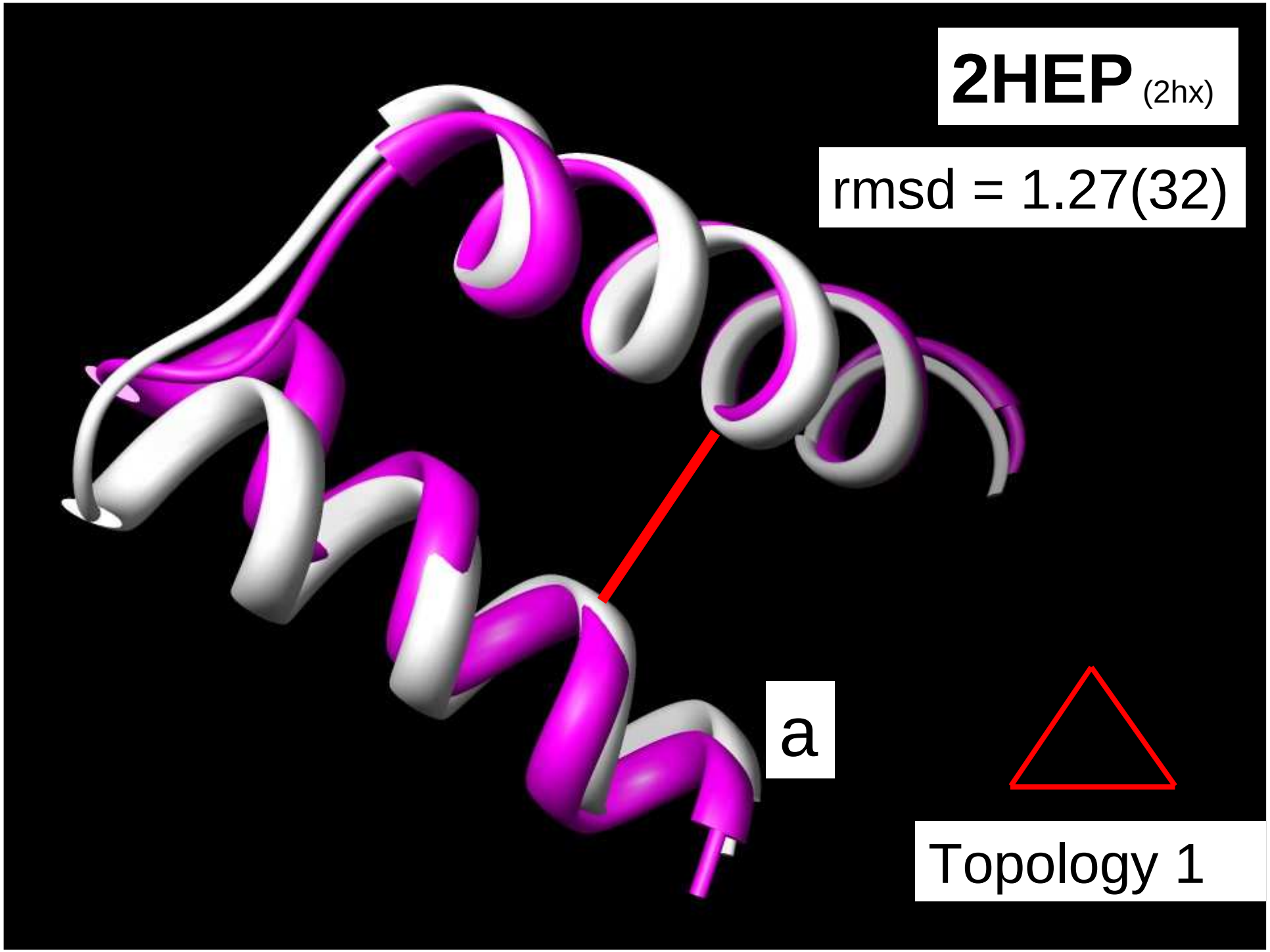
Ensemble of structures generated, ranked by RgH, clustered by mutual RMSD, lowest RgH structure kept per cluster

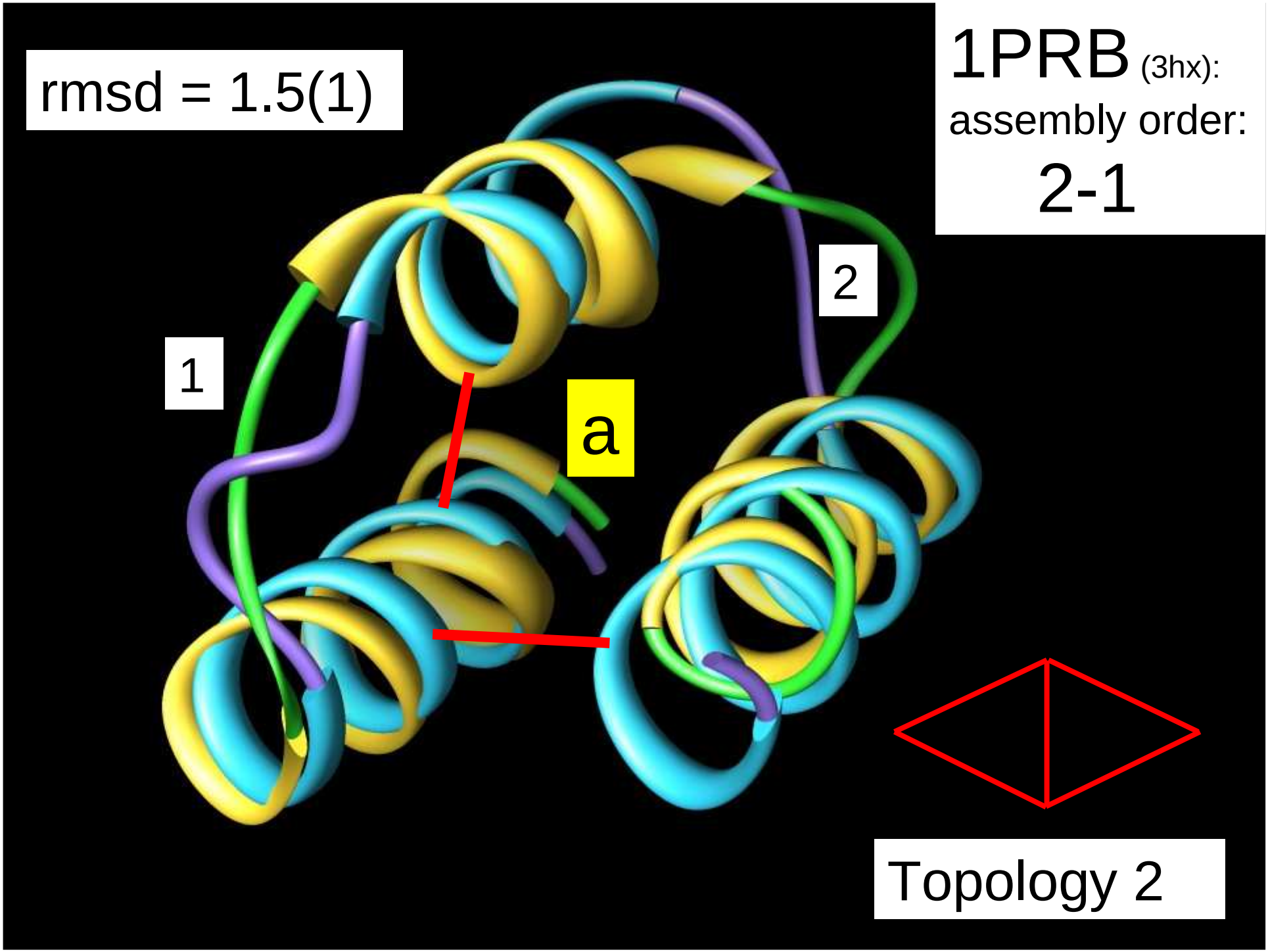
Cluster cutoff heuristic: (n-1)Ang, n = #of helices in assembly

- starting with long loops, we would need to keep less compact conformations (in addition to compact ones) in order to ensure the native conformation is covered.

- starting with long loops, we would need to keep less compact conformations (in addition to compact ones) in order to ensure the native conformation is covered.
- by adding long loop closure at the later steps, we have a more limited conformation space to explore due of excluded volume effects (i.e. steric constraints with the pre-assembled parts).

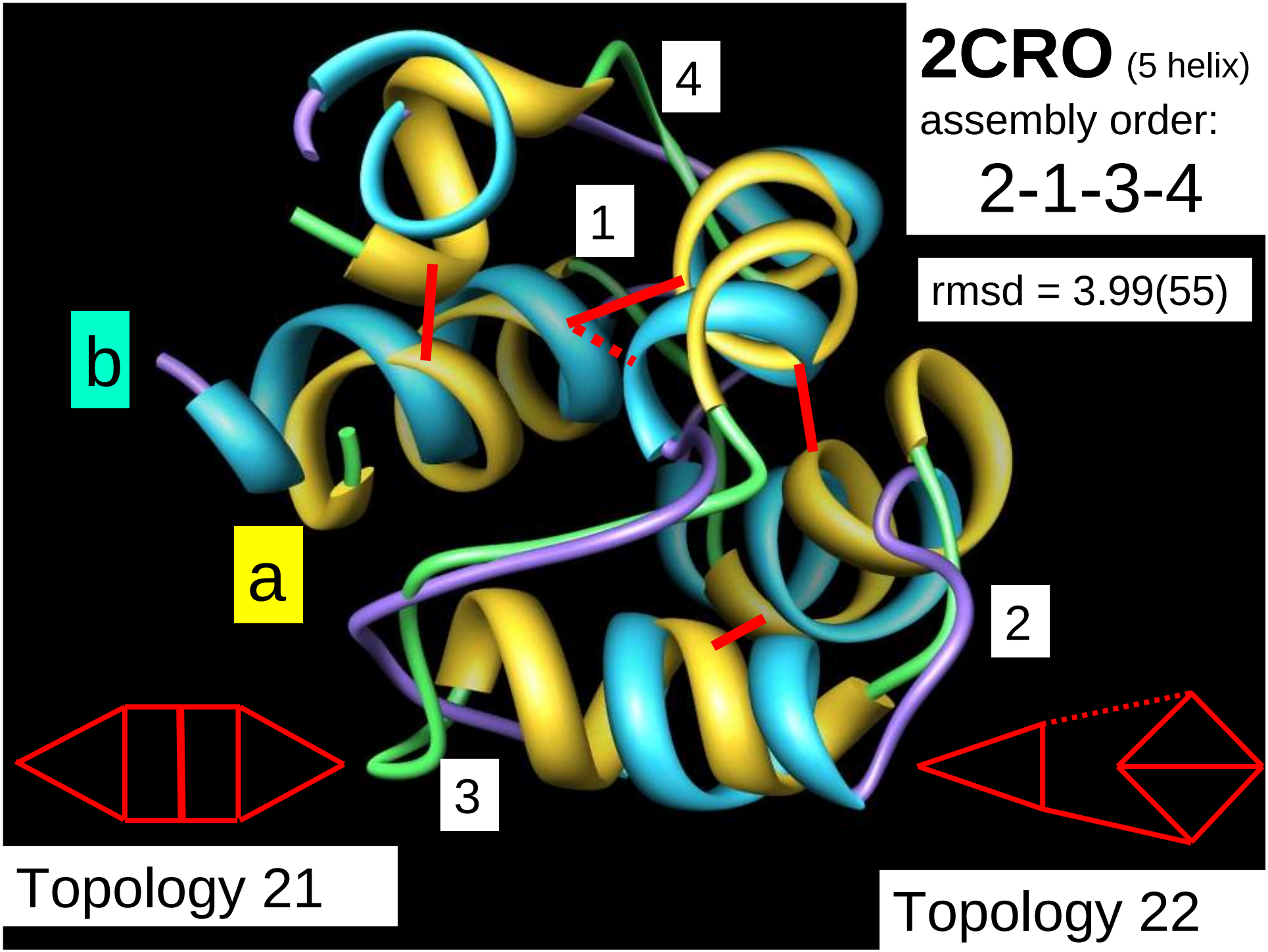
- starting with long loops, we would need to keep less compact conformations (in addition to compact ones) in order to ensure the native conformation is covered.
- by adding long loop closure at the later steps, we have a more limited conformation space to explore due of excluded volume effects (i.e. steric constraints with the pre-assembled parts).
- Amber force field energy minimization is done after loop closure for better sterics (30-60 sd/cg steps)


2HEP (2hx)
rmsd = 1.27(32)
a
Topology 1


rmsd = 1.5(1)
1PRB (3hx):
assembly order:
2-1
2
1
a
Topology 2

Larger proteins

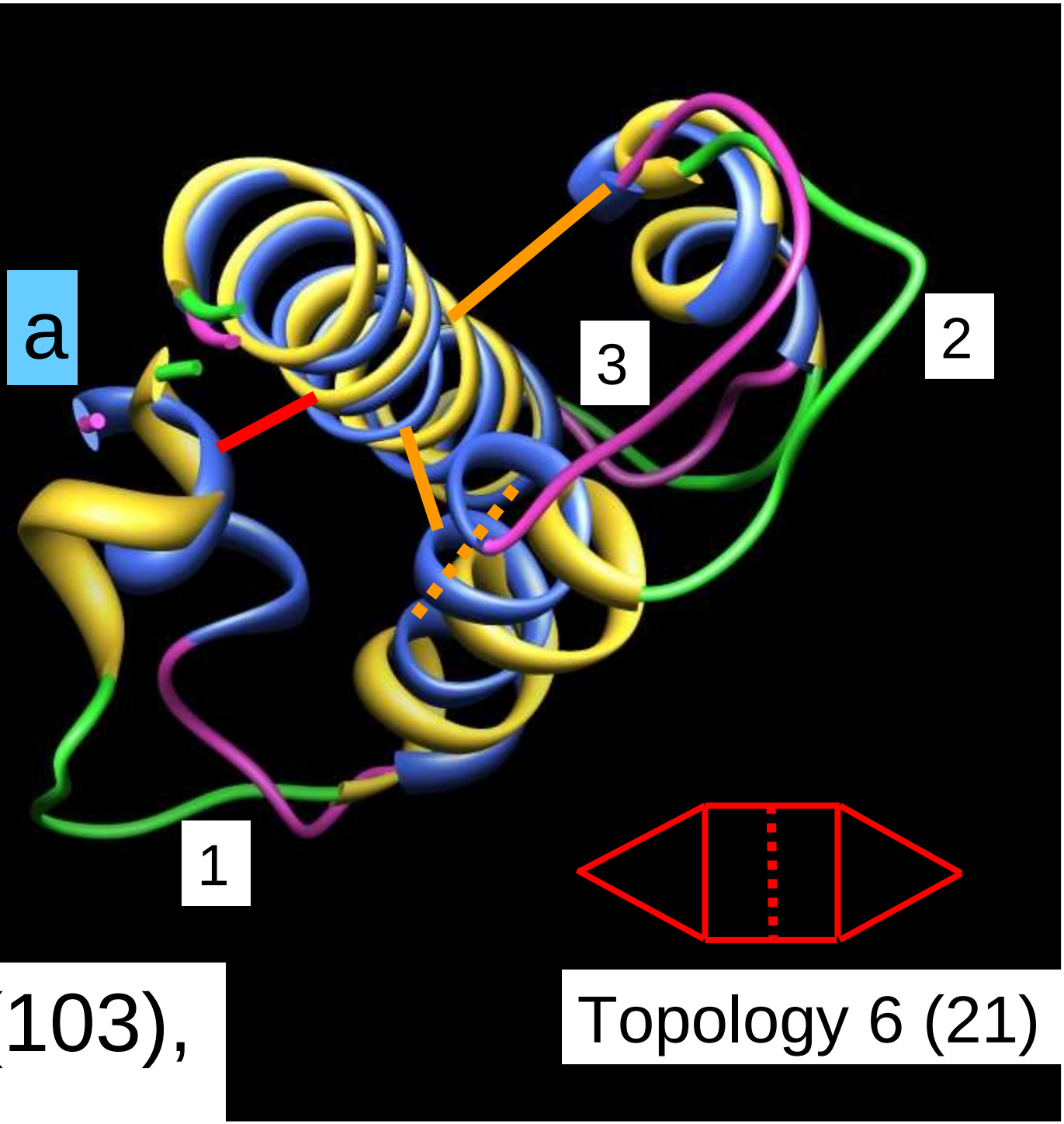
- subtle connectivity topology errors may exist, although overall RMSD is still good


2CRO (5 helix)
assembly order:
2-1-3-4
rmsd = 3.99(55)
4
1
b
a
2
3
Topology 21
Topology 22

Disulfide bridged proteins

- RgH scoring not discriminating: native not among few best structures
- Imposing known disulfide bonds introduces sufficient restrictions (at least in the 5 proteins we attempted) that still allowed us to sample near-native structures
- TLC applicable to S2 bridges (but not included in current implementation); used amber9 with restraints to close bridges for these studies

1C5A (4hx)
assembly order:
3-2-1



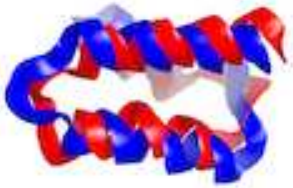
RMSD = 2(103),

Topology 6 (21)

Helical protein gallery: **native** vs. lowest rmsd **model**

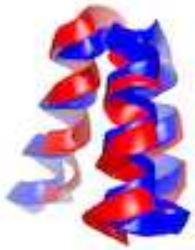
2HEP
1.27(32)

1RPO
1.52(3)



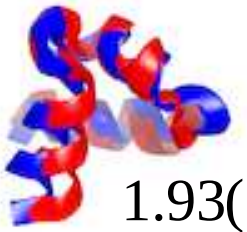
1BDD
1.58(24)

1DV0
1.92(19)



1GAB
1.67(21)

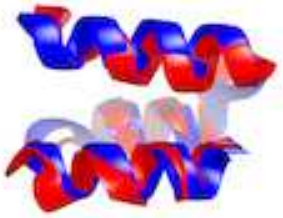
1GVD
1.51(13)



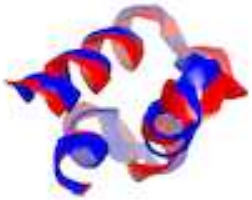
1.93(6)

1IDY

1MBH
1.80(20)

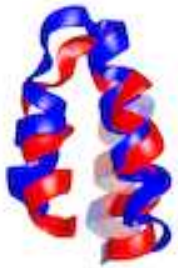


1PRB
1.50(1)



1PRV
2.03(13)

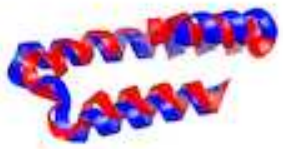
1G2H
1.80(1)



1X9B
2.21(2)

1ENH
2.50(34)

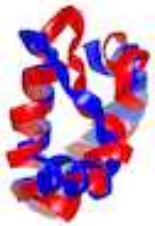
1FEX
2.67(10)



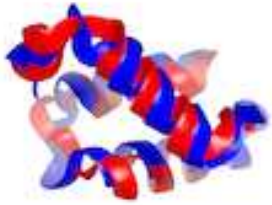
1LRE
2.96(19)

2A3D
2.29(30)

1I6Z
2.91(8)

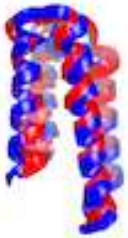


1EIJ
4.54(22)

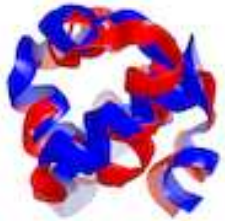


2EZH
3.21(47)

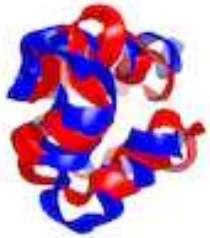
1POU
4.22(42)



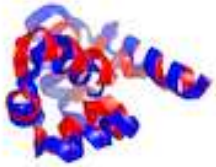
2MHR
2.14(8)



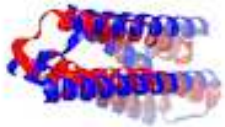
1R69
4.00(43)



2CRO
4.01(12)



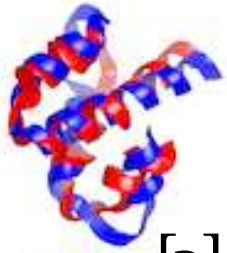
2ICP
5.10(15)



1LPE
4.59(23)

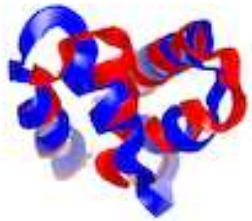
1HP8 [3]
2.45(30)

1ERY [2]
1.69(9)



1C5A [3]
2.49(17)

1GH1 [2]
3.21(2)



1J0T [2]
2.73(28)