Protein Loop Kinematics

A protein loop (or fragment) L is a sequence of $p > 3$ consecutive residues in a protein P, such that none of the two termini of L is also a terminus of P. We number the residues of L from 1 to p, starting at the N terminus. We model the kinematics of the backbone of L as a serial linkage whose DOFs are the $n = 2p$ dihedral angles ϕ_i and ψ_i around the bonds N–Cα and Cα–C, in residues $i = 1, ..., p$. The rest of the protein, denoted by $P\backslash L$, is assumed rigid. We let L_B denote the backbone of L. It includes the Cβ and O atoms respectively bonded to the Cα and C atoms in the backbone, as their positions are uniquely determined by the dihedral angles ϕ_i and ψ_i.

To simplify our presentation, we will use the above model throughout. However, our methods are more general. They allow a loop to start at either an N or Cα atom and end at either a Cα or C atom. Furthermore, they do not require all ϕ and ψ angles in L_B to be variable. Some may have fixed input values. We only require the loop to contain at least three amino-acids, consecutive or not, with variable ϕ and ψ angles. This condition is needed for the application of the analytical IK method described in [Coutsias *et al.*, 2004].

We attach a Cartesian coordinate frame Ω_1 to the N terminus of L and another frame Ω_2 to its C terminus. When L_B is connected to the rest of the protein, i.e., when it adopts a *closed* conformation, the pose (position and orientation) of Ω_2 relative to Ω_1 is fixed. We denote this pose by Π_g. However, if we arbitrarily pick the values of ϕ_i and ψ_i, $i = 1$ to p, then in general we get an *open* configuration of L_B, where the pose of Ω_2 differs from Π_g. The set $\mathbf{Q}$ of all open and closed conformations of L_B is a manifold of dimensionality $n = 2p$. The subset $\mathbf{Q}_{\text{closed}}$ of closed conformations is a submanifold of $\mathbf{Q}$ of dimensionality $n - 6$ for almost any Π_g; it is sometimes called the "self-motion manifold" of L_B. When $n = 6$, $\mathbf{Q}_{\text{closed}}$ has zero dimensionality; it then contains at most 16 isolated closed conformations [Coutsias *et al.*, 2004, Liu *et al.*, 2006]. Let $\Pi(q)$ denote the pose of Ω_2 relative to Ω_1 when the conformation of L_B is $q \in \mathbf{Q}$. The function Π and its inverse Π^{-1} are the "forward" and "inverse" kinematics map of L_B, respectively. We have: $q \in \mathbf{Q}_{\text{closed}} \Leftrightarrow \Pi(q) = \Pi_g$.

A conformation of L_B is *clash-free* if and only if no two atoms, one in L_B, the other in L_B or $P\backslash L$, are such that their centers are closer than ε times the sum of their van de Waals radii, where ε is a constant in $(0, 1)$. In our software, ε is an adjustable parameter, usually set to 0.75, which approximately corresponds to the distance where the van der Waals potential associated with two atoms begins increasing steeply. We denote the set of closed clash-

free conformations of L_B by $\mathbf{Q}_{\text{closed}}^{\text{free}}$. It has the same dimensionality as $\mathbf{Q}_{\text{closed}}$, but its volume is usually a small fraction of that of $\mathbf{Q}_{\text{closed}}$.

References

Coutsias, E.A., Soek, C., Jacobson, M.P., and Dill, K.A. A kinematic view of loop closure, *Journal of Computational Chemistry*, **25**:510-528, 2004.

Liu, G., Milgram, J., Dhanik, A., and Latombe, J.C. On the inverse kinematics of a fragment of protein backbone. *Proc. 10th Symp. on Advances in Robot Kinematics* (ARK), Ljubljana, Slovenia, June 2006.