SUPPLEMENTARY INFORMATION FOR CHAINTWEAK: SAMPLING FROM THE NEIGHBOURHOOD OF A PROTEIN CONFORMATION

ROHIT SINGH and BONNIE BERGER*†

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge MA 02139
E-mail: {rsingh, bab}@mit.edu

This document provides some supplementary information for:
Singh R. and Berger B. *ChainTweak: Sampling From the Neighbourhood Of a Protein Conformation* Proceedings of the Pacific Symposium of Biology, 2005.
This document and related material can always be found at `http://theory.csail.mit.edu/chaintweak`

1. Formal Problem Definition

*This section formally defines the neighbourhood sampling problem (*NeighSample*) and the related* CTprob *problem.*

The neighbourhood sampling problem can be formally stated as follows:
Problem 1 NeighSample *Given a base conformation $\mathcal{C}_0$, find a set of N conformations $S = \{\mathcal{C}_i\}$ such that the RMSDs of these conformations w.r.t. the base roughly follows some desired distribution $\mathcal{D}$.* For example, we might want that half of the conformations be 0–2Å RMSD from the base while the rest be 2–4Å RMSD from the base. Choosing an appropriate D lets us create a problem instance suitable for the application at hand.

We solve the NeighSample problem as follows:

S1. given $\mathcal{C}_0$, generate r conformations $\mathcal{C}_1^T \ldots \mathcal{C}_r^T$ by randomly sampling new values for dihedral angles at the N and C-terminal residues; r is a small integer (e.g., 10). The sampling can be biased, using a Ramachandran Plot[1], to pick desirable phi/psi values.

S2. for each $\mathcal{C}_i^T$, generate N/r conformations by using ChainTweak. Output these. $\mathcal{D}$ influences the choice of two parameters in the ChainTweak algorithm: K, the number of iterations, and `Filter`, the pruning policy at the end of each iteration (see Algorithm 1).

The motivation behind varying the positions of terminal residues *a-priori* is that ChainTweak does not move the terminal residues. Recall that arbitrarily modifying dihedral angles near the ends does not result in large deviations for the entire structure (Fig 1b). In cases where the terminal residues do need to remain fixed, e.g. loop modeling, or when only a part of structure should be modified, Step

*Corresponding author
†Also in the MIT Dept. of Mathematics

1 may be skipped. Formally, ChainTweak solves the CTPROB problem. In the following problem statement, a backbone conformation $\mathcal{C}(\mathbf{v_0}, \Theta)$ is defined by the three initial atom positions $\mathbf{v_0} = \langle v_0, v_1, v_2 \rangle$ and a sequence of dihedral angles $\Theta = (\theta_1, \ldots, \theta_{n-2})$.

Problem 2 CTPROB *Given 3 initial atom positions* $\mathbf{v_0}$ *and a starting backbone conformation* $\Theta^0 = (\theta_1^0, \ldots, \theta_{n-2}^0)$ *let* $\mathbf{v_{n-2}} = \langle v_{n-2}, v_{n-1}, v_n \rangle$ *be the last 3 points in* $\mathcal{C}_0 = \mathcal{C}(\mathbf{v_0}, \Theta^0)$. *Find N other sets of dihedral angles such that for each such set* Θ *the last 3 points of the conformation* $\mathcal{C}(\mathbf{v_0}, \Theta)$ *are still* $\mathbf{v_{n-2}}$.

Observe that this problem formulation has two useful properties. First, solutions to CTPROB can be *recursively* found: a chain can be broken into two sub-chains and alternative conformations found for each of the two sub-chains can be joined to produce an alternative conformation for the entire chain. Second, solutions can be iteratively found: if $\mathcal{C}_1 = \mathcal{C}(\mathbf{v_0}, \Theta^1)$ is an alternative conformation for $\mathcal{C}_0 = \mathcal{C}(\mathbf{v_0}, \Theta^0)$ and $\mathcal{C}_2 = \mathcal{C}(\mathbf{v_0}, \Theta^2)$ is an alternative conformation for $\mathcal{C}_1$, then $\mathcal{C}_2$ is an alternative conformation for the original conformation $\mathcal{C}_0$.

2. Formal Description of Algorithms

This section formally defines the algorithms used by ChainTweak.

Algorithm 1 ChainTweak: solve the CTprob problem, producing N new conformations from the given end points and starting conformation

Input: chain $P = \langle v_0, \ldots, v_n \rangle$ of n bonds
N, the number of new conformations desired
bond lengths $\mathbf{l} = \langle l_1, \ldots, l_{n-2} \rangle$, l_i is defined by v_i, v_{i+1}
bond angles $\boldsymbol{\beta} = \langle \beta_1, \ldots, \beta_{n-2} \rangle$, β_i is defined by v_i, v_{i+1}, v_{i+2}
Output: $S = \{P' \mid P'$ is an alternative conformation, as defined in SlideWin$\}$
Parameters: K: the number of iterations
`Filter`: a policy for choosing appropriate conformations

$S \longleftarrow \{P\}$
for $i = 1 \ldots K$ **do**
  $S' \longleftarrow \Phi$
  for $j = 1 \ldots N$ **do**
    pick an element X from the set S randomly
    $X' \longleftarrow$SlideWin$(X, \mathbf{l}, \boldsymbol{\beta})$
    $S' \longleftarrow S' \cup \{X'\}$
  $S \longleftarrow$`Filter`(S')
return S

Algorithm 2 SlideWin: solve the CTprob problem, producing one new conformation from the given end points and starting conformation

Input: chain $P = \langle v_0, \ldots, v_n \rangle$ of n bonds
bond lengths $\mathbf{l} = \langle l_1, \ldots, l_{n-2} \rangle$, l_i is defined by v_i, v_{i+1}
bond angles $\boldsymbol{\beta} = \langle \beta_1, \ldots, \beta_{n-2} \rangle$, β_i is defined by v_i, v_{i+1}, v_{i+2}
Output: $P' = \langle v'_0, \ldots, v'_n \rangle$
where $\langle v_0, v_1, v_2 \rangle = \langle v'_0, v'_1, v'_2 \rangle$ and $\langle v_{n-2}, v_{n-1}, v_n \rangle = \langle v'_{n-2}, v'_{n-1}, v'_n \rangle$
LoopClsr6: Input: the fixed end points and bond lengths/angles of a 6-DOF chain
Output: $S = \{\Theta \mid$ the dihedral angles Θ specify an alternative conformation$\}$

$P' \longleftarrow P$
for $i = 0, 3, 6, \ldots, 3\lfloor n/3 \rfloor$ **do**
  $S_\Theta \longleftarrow$LoopClsr6$(\langle P'[i], P'[i+1], P'[i+2] \rangle, \langle v_{i+6}, v_{i+7}, v_{i+8} \rangle, \langle l_{i+1}, \ldots, l_{i+7} \rangle, \langle \beta_{i+1}, \ldots, \beta_{i+7} \rangle)$
  if S_Θ is non-empty **then**
    pick an element Θ of S_Θ randomly, biasing the choice as per the phi-psi preference map for residue $1 + \lfloor i/3 \rfloor$
    calculate $\langle x_0, \ldots, x_8 \rangle = \mathcal{C}(\langle P'[i], P'[i+1], P'[i+2] \rangle, \Theta)$
    $P'[i+3], \ldots, P'[i+8] = x_3, \ldots, x_8$
return P'

Algorithm 3 LoopClsr6: solve the loop closure problem for chains with 9 atoms (and 8 bonds)

Input: $\langle v_0, v_1, v_2 \rangle$ are fixed positions of the first 3 points
$\langle v_6, v_7, v_8 \rangle$ are fixed positions of the last 3 points
$\langle l_1, \ldots, l_6 \rangle$ is the list of bond lengths. l_i is the length of the link $v_i v_{i+1}$
$\langle \beta_1, \ldots, \beta_6 \rangle$ is the list of bond angles. β_i is defined by the points v_i, v_{i+1}, v_{i+2}
Output: $S = \{\Theta \mid \Theta = \langle \theta_1, \ldots, \theta_6 \rangle$ and the chain $\langle \mathbf{v_0}, \Theta \rangle$ has $\mathbf{v_6}$ as its last 3 points $\}$
where $\mathbf{v_0} = \langle v_0, v_1, v_2 \rangle$, $\mathbf{v_6} = \langle v_6, v_7, v_8 \rangle$, and $|S| \leq 16$.

if omega angles can take arbitrary values **then**
  refer to Manocha *et al.*[2]
if omega angles should have same value as before or should be 180° **then**
  refer to Coutsias *et al.*[4]

3. Numerical Accuracy In ChainTweak

ChainTweak has high numerical accuracy. When omega angles are restricted to particular values, the error in the final angle values is less than 10^{-4} degrees, averaged over 10000 conformations. Similarly, every call to the loop closure routine results in a small numerical error in the placement of the end-residues inside the sliding window. Left unchecked, these errors accumulate, resulting in significant errors at the extreme ends of the chain. Our couour implementation of ChainTweak avoids error accumulation across multiple calls to loop closure routines. Because the ends are not supposed to move, we just recalibrate the (slightly errorneous) end-positions against the original end-positions. As a result, the deviation of atom positions in the terminal residues is negligible: avg error = 0.001Å.

References

1. Branden C, Tooze J. *Introduction to Protein Structure*. Garland Pub, NY, 1991
2. Manocha D *et al. Conformational analysis....* Comp App of Bio Sci, 11(1):71-86, 1995
3. Wedemeyer WJ, Scheraga HA. *Exact Analytical...* J Comp Chem, 20:819-844, 1999
4. Coutsias EA *et. al. A kinematic view....* J Comp Chem 25(4):510-28, 2004