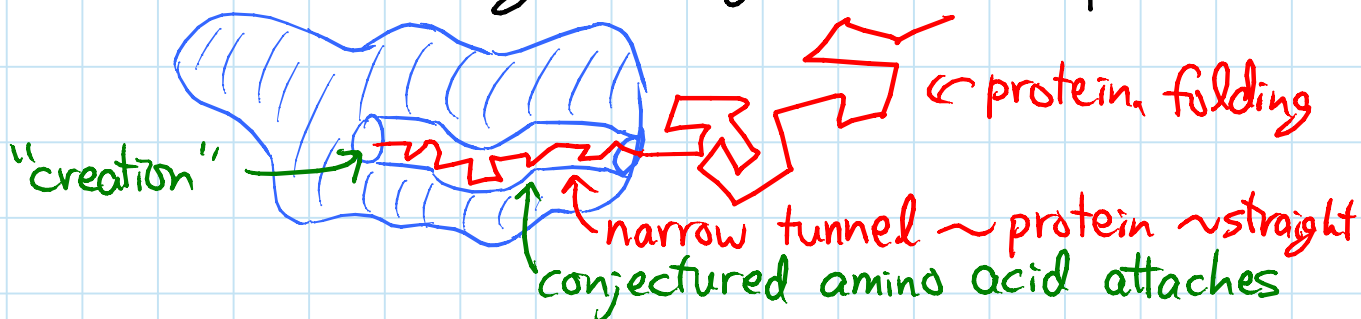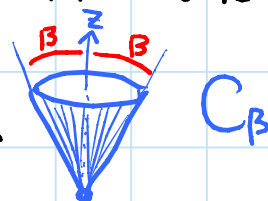# Producible protein (fixed-angle) chains:

[Demaine, Langerman, O'Rourke 2003/6]
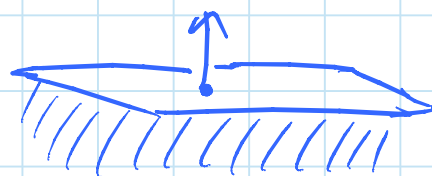
Ribosome = "machine" built from proteins & RNA
translating messenger RNA into proteins

"creation"

protein folding

narrow tunnel ~ protein ~ straight

conjectured amino acid attaches

β-producible chain = simple geometric model of
chains & configurations resulting from ribosome
- cone C_β of half-angle β
- chain produced in cone, link by link
- latest link passes through cone apex
- when latest vertex v_i reaches cone apex, next link (v_i, v_{i+1}) is instantly created in cone & v_i can never re-enter cone

C_β

Reality: $\beta = 90°$ (halfspace)
is the closest model
(though model is rather local—
protein might extend around ribosome
& violate our abstraction)

Angle limitations in β-producible chains:

- when link (v_i, v_{i+1}) passes through cone apex
  (v_i outside cone, v_{i+1} inside cone),
  v_i is in <u>complementary cone</u> B_β:
  - if $\beta \leq 90°$ then B_β = reflection of C_β thru xy
  - if $\beta \geq 90°$ then B_β = complement/exterior of C_β
    & v_{i+1} is actually in reflection of B_β thru xy

$\Rightarrow v_i$ & v_{i+1} on opposite sides of double-cone
of half-angle $\min \{\beta, 180° - \beta\}$

- when (v_{i+1}, v_{i+2}) created inside C_β,
  both (v_i, v_{i+1}) and (v_{i+1}, v_{i+2}) are in the double-cone

$\Rightarrow$ max turn angle $\alpha \leq 2 \min \{\beta, 180° - \beta\}$

i.e. $\alpha/2 \leq \beta \leq 180° - \alpha/2$

<u>($\leq\alpha$)-chain</u> = chain of max. turn angle $\leq \alpha$

Canonical configuration for $(\leq\alpha)$-chains:

- put v_0 at origin $(0,0,0)$
- put v_{i+1} on cone $C_{\alpha/2}$ centered at v_i
- v_1 chosen to maximize x coordinate
- v_{i+1} chosen to get correct turn angle at v_i:
  - view on sphere centered at v_i & radius $\alpha/2$
  - $C_{\alpha/2}$ intersects along circle around north pole
  - turn-angle cone intersects along tilted circle of radius τ_i
  - intersections overlap (at 1 or 2 pts.) because center of turn-angle circle is on $C_{\alpha/2}$ circle & $\tau_i \leq \alpha$
  - take <u>counterclockwise-most</u> intersection for v_{i+1}
      ↳relative to origin
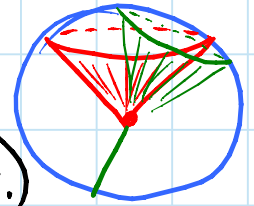
- kind of spiral
- $\sim$ similar to nature's helix
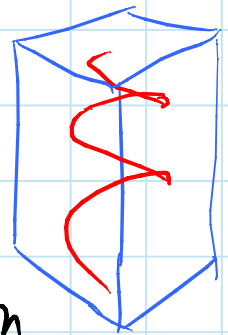
- contained in $C_{\alpha/2}$ cone: by induction

  $C_{\alpha/2}$ ← rest of chain (by induction)
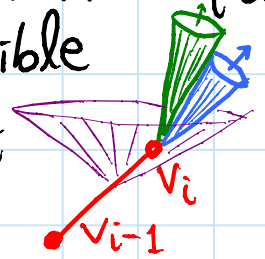 ← first link (by construction)
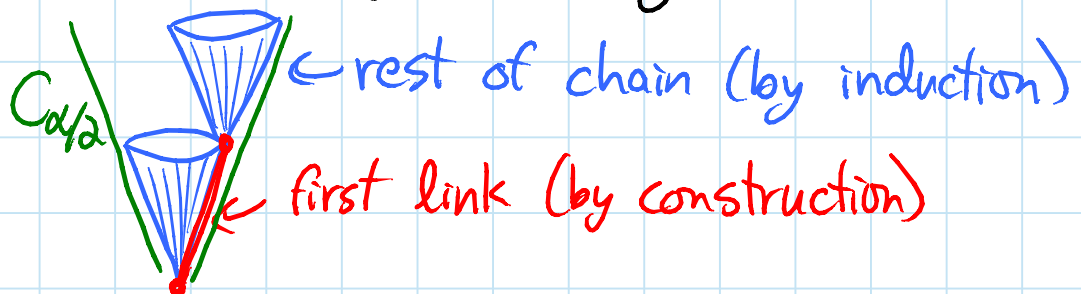
  - in fact, strictly inside cone $C_{\alpha/2}$ except for first link because (v_0, v_1) & (v_1, v_2) not parallel

Canonicalizing $(\geq \alpha)$-producible $(\leq \alpha)$-chains:

- **main idea**: play production movie backwards
⇒ as links enter the cone, they disappear
- **maintain** these links in canonical configuration, translated to start at last existing vertex v_i & rotated to make <u>cone</u> as vertical as possible
   while satisfying <u>turn angle</u> at v_i
    - viewed on sphere centered at v_i:
    put <u>canonical cone</u> axis
    $2\tau_i$ up from previous edge direction
    toward north pole (maxing out at north pole)
    ⇒ canonical configuration is in $C_{\beta (\geq \alpha)}$
    because (v_{i-1}, v_i) is too (by production)
    - if (v_{i-1}, v_i) is vertical, then
    orientation of first link is not determined
        - choices for smaller & larger times may differ
        - freeze movie & continuously spin (v_{i-1}, v_i)
        to switch from previous choice to next
    - when v_i reaches cone apex, need to extend
    canonical configuration & maintain invariant
        - spin (v_{i-1}, v_i) to make (v_i, v_{i+1}) as vertical
        as possible ⇒ new canon. config. rotation
        - spin (v_i, v_{i+1}) to bring (v_i, v_{i+1}) into
        canonical configuration
        - note: already canonical ⇒ rigid   □

What is producible?

- α-canonical configuration is β-producible
  for $\alpha/2 \leq \beta \leq 180° - \alpha/2$ (full range)
  - keep canonical configuration in
    complementary cone B_β
  $\Rightarrow$ produces "rigidly" (no spinning required)
- $(\leq\alpha)$-chain $(\geq\alpha)$-producible
          $\Rightarrow$ β-producible for $\alpha/2 \leq \beta \leq 180° - \alpha/2$
  - β-produce α-canonical configuration
  - reverse canonicalization procedure
    far away from production cone C_β
- flat states of $(\leq\alpha)$-chains are β-producible
  for $\alpha \leq \beta \leq 90°$
  - imagine moving cone instead of chain
  - create <u>next link</u>
    in vertical plane
  - slide up to plane of
    flat configuration with cone just touching plane
  - repeat
$\Rightarrow$ flat-state connected
  - canonicalize both, combine motions
$\Rightarrow$ for $(\leq\alpha)$-chains & $\alpha \leq \beta \leq 90°$,
  configuration is flattenable $\Leftrightarrow$ it is β-producible
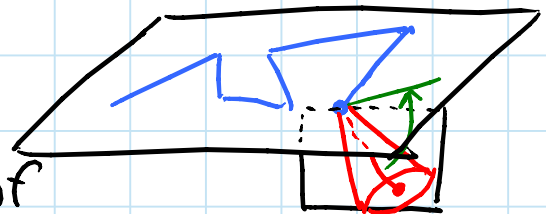
HP model of protein folding energetics: [Ken Dill 1990]
simple combinatorial model capturing an often
prominant force: hydrophobic bonds (avoid surrounding water)
— protein = chain of (H) or (P) nodes
   hydrophobic↵   ↳hydrophilic/polar
— folding = embedding on square grid in 2D or 3D
   (also several variations, avoiding parity issue, e.g.,
   equilateral Δ in 2D & cannonball packing in 3D)
— score = # pairs of (H) nodes adjacent in lattice
   (usually ignoring edges of the chain — forced)
— optimal folding = maximum score

Finding optimal folding is NP-hard:
   — in 3D   [Berger & Leighton 1998]
   — in 2D   [Crescenz, Goldman, Papadimitriou, Piccalboni, Yannakakis 1998]
   — various O(1)-approximations
   [Hart & Istrail 1996; Mauri et al. 1999; Newman 2002;
   Agarwala, Batzoglou, Dancik, Decatur, Farach, Hannenhali,
   Muthukrishnan, Skiena 1997]

OPEN: $(1-\varepsilon)$-approximation?   APX-hard?

HP model approximation: (square grid)
— upper bound: $2 \cdot \min\{\#\text{ even H's}, \#\text{ odd H's}\}$
  — every bond is between odd & even H
  — ≤ 2 bonds per H, but double counting
  $\Rightarrow \leq 2$ bonds per even H, and ditto for odd H
— assume $\#$ even H's $= \#$ odd H's (excess $\to$ P's)

Best approximation: $\frac{1}{3}$ [Newman 2002]
  — works even for closed chains (harder)
  — find breakpoint such that
    — proceeding clockwise, always at least
      as many odd H's as even H's
    — proceeding counterclockwise, always at least
      as many even H's as odd H's
  — hang chain down from breakpoint edge
  — match up even H's of left (ccw) half
            & odd H's of right (cw) half
  — get 2 bonds for every 3 H's
  — but ignoring half the H's
  $\Rightarrow 1$ bond per 3 H's
  — upper bound $= \# H's$
  $\Rightarrow \frac{1}{3}$ -approx.   □

Unique optimal foldings: [Aichholzer, Bremner, Demaine, Meijer, Sacristán, Soss 2003]

exist in 2D square grid
- for all even n for closed chains
- for doubly even n for open chains

n = 20
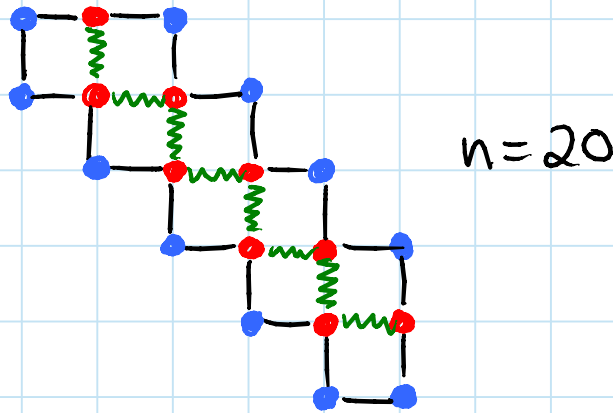
OPEN : other grids than square? (presumably)

OPEN : optimal folding NP-hard among uniquely optimal proteins? (probably)

OPEN : design protein with unique optimal folding with approximately optimal shape?
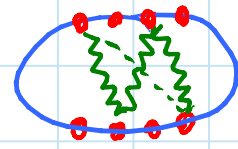- potentially polynomial!

Proof of closed-chain unique optimal foldings:

- Zigzag folding achieves score $h-1$

  (# H nodes)

- optimal always $\leq h$ (≤ 2 contacts per H, double counting)

- here, optimal $\leq h-1$:
  - ≥ 4 nonconsecutive edges on bounding box, one against each wall
  - ≤ 2 are P–P edges
  - $\Rightarrow \geq 2$ have H nodes
  - these H nodes lose ≥ 1 contact
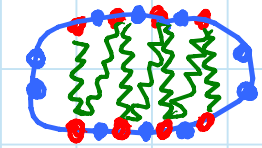 - $\Rightarrow$ lose 1 contact (double counting)

- parity: $h/2$ "even" H nodes, $h/2$ "odd" H nodes
- only bonds between even & odd H nodes
- no contacts external to closed chain
  - P–P edges must be on bounding box
  - partitions odd & even H nodes
- graph of contacts is acyclic
  - bipartite & "inside"
- $\Rightarrow$ graph is a path
- path is unique up to reflection
- decomposes into squares:

$\Rightarrow$ zigzag