Algorithmica manuscript No.
(will be inserted by the editor)

Geometric Restrictions on Producible Polygonal Protein Chains

Erik D. Demaine[1*], **Stefan Langerman**[2**], **Joseph O'Rourke**[3***]

[1] MIT Computer Science and Artificial Intelligence Laboratory,
32 Vassar Street, Cambridge, MA 02139, USA, e-mail: `edemaine@mit.edu`
[2] Université Libre de Bruxelles, Département d'informatique,
ULB CP212, Bruxelles, Belgium, e-mail: `Stefan.Langerman@ulb.ac.be`
[3] Department of Computer Science, Smith College,
Northampton, MA 01063, USA, e-mail: `orourke@cs.smith.edu`

The date of receipt and acceptance will be inserted by the editor

Abstract Fixed-angle polygonal chains in 3D serve as an interesting model of protein backbones. Here we consider such chains produced inside a "machine" modeled crudely as a cone, and examine the constraints this model places on the producible chains. We call this notion *producible*, and prove as our main result that a chain whose maximum turn angle is α is producible in a cone of half-angle $\geq \alpha$ if and only if the chain is flattenable, that is, the chain can be reconfigured without self-intersection to lie flat in a plane. This result establishes that two seemingly disparate classes of chains are in fact identical. Along the way, we discover that all producible configurations of a chain can be moved to a canonical configuration resembling a helix. One consequence is an algorithm that reconfigures between any two flat states of a "nonacute chain" in $O(n)$ "moves," improving the $O(n^2)$-move algorithm in [ADD$^+$02].

Finally, we prove that the producible chains are rare in the following technical sense. A random chain of n links is defined by drawing the lengths and angles from any "regular" (e.g., uniform) distribution on any subset of the possible values. A random configuration of a chain embeds into $\mathbb{R}^3$ by in addition drawing the dihedral angles from any regular distribution. If a class of chains has a locked configuration (and no nontrivial class is known to avoid locked configurations), then the probability that a random

* Supported by NSF CAREER award CCF-0347776 and DOE grant DE-FG02-04ER25647.
** Chercheur qualifié du FNRS.
*** Supported by NSF Distinguished Teaching Scholars award DUE-0123154.

configuration of a random chain is producible approaches zero geometrically as $n \to \infty$.

1 Introduction

The backbone of a protein molecule may be modeled as a 3D polygonal chain, with joints representing residues and fixed-length links (edges) representing bonds. The joints are not universal; rather successive bonds form nearly fixed angles in space. The motions at the joints are then called *dihedral* motions. The study of such *fixed-angle* chains was initiated in [ST00] and continued in [ADM$^+$02] and [ADD$^+$02]. These papers identified *flat states* of a chain—embeddings into a plane without self-intersection—as geometrically interesting. A chain that can reconfigure in $\mathbb{R}^3$ via dihedral motions between any two of its flat states is called *flat-state connected*. A chain that has a flat state but is in a configuration that cannot reach that state (via dihedral motions, without self-intersection) is called *unflattenable* or simply *locked*.[1]

We look here at a particularly simple but natural constraint on the "production" of a fixed-angle chain. Our inspiration derives from the ribosome, which is the "machine" that creates protein chains in biological cells. Fig. 1 shows a schematic cross section of a ribosome and its exit tunnel, based on a model developed by Nissen et al. [NHB$^+$00]. We consider a very simple geometric model that roughly captures the exit point x of the ribosome: the chain is produced inside a cone of some half-angle β, with the chain emerging through the cone's apex. See Fig. 2. This constraint immediately implies that the maximum turn angle α in the produced chain is at most 2β. We consider the somewhat stronger condition that $\alpha \leq \beta$. These conditions are consistent with our analogy to the ribosome, where the cone is roughly a half plane (half-angle $\beta = 90°$) and the chain has obtuse angles around $110°$ (turn angle $\alpha = 70°$).

We show in Section 3 that this simple constraint guarantees that all producible chains are flattenable and furthermore mutually reachable. There are several interesting aspects to this result. First, we are naturally led in our proof to a canonical form, called α-CCC, which bears a resemblance to the helical form preferred by many proteins. Second, we show in Section 5 that long "random" chains are locked with probability approaching 1, implying that producible protein chains are rather special. Third, we show in Section 4 that, if we strengthen the production model to allow producing chain turn angles of more than 2β, then locked chains can be produced. This example shows the importance of our condition that $\alpha \leq \beta$ (or a similar condition such as $\alpha \leq 2\beta$).

[1] In fact, this definition is slightly more specific than the usual notion of "locked," which says that there are two arbitrary configurations of the linkage that are mutually unreachable.

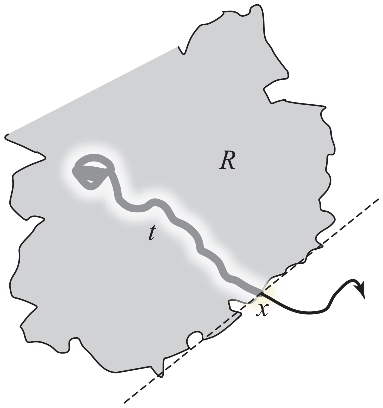

Fig. 1 The ribosome R in cross-section. The protein is created in tunnel t and emerges at x.

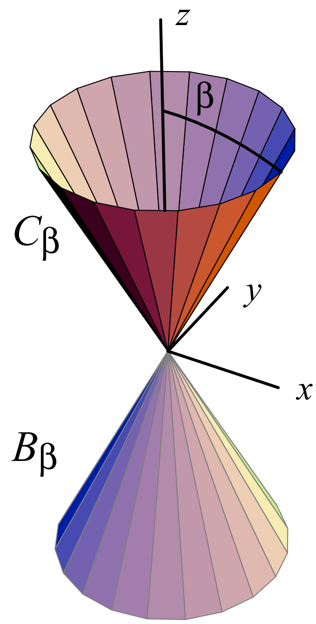

Fig. 2 The chain is produced in cone C_β, and emerges at the origin into the complementary cone B_β below the xy-plane.

2 Definitions

2.1 Chains and Motions

The fixed-angle polygonal chain P has $n+1$ vertices $V = \langle v_0, \ldots, v_n \rangle$ and is specified by the fixed turn angle θ_i at each vertex v_i, $i = 1, \ldots, n-1$, and by the edge length d_i between v_i and v_{i+1}, $i = 0, \ldots, n-1$. When all angles $\theta_i \leq \alpha$ for some $0 < \alpha < \pi$, P is called a $(\leq \alpha)$-*chain*.[2] We write $P[i,j]$, $i \leq j$, for the polygonal subchain composed of vertices $v_i, \ldots, v_j$.

A *configuration* $Q = \langle q_0, \ldots, q_n \rangle$ of the chain P (see Fig. 3) is an embedding of P into $\mathbb{R}^3$, i.e., a mapping of each vertex v_i to a point $q_i \in \mathbb{R}^3$, satisfying the constraints that the angle between vectors $q_{i-1}q_i$ and q_iq_{i+1} is θ_i, and the distance between q_i and q_{i+1} is d_i. The points q_i and q_{i+1} are connected by a straight line segment e_i. Thus, a configuration can be specified by the position of e_0 and dihedral angles δ_i, $i = 1, \ldots, n-2$, where δ_i is the angle between planes $e_{i-1}e_i$ and e_ie_{i+1}. The configuration is *simple* if no two nonadjacent segments intersect.

[2] Other work [ADM$^+$02, ADD$^+$02] focuses on the angle between adjacent edges, which for us is $\pi - \alpha$. Thus "nonacute chains" in that work corresponds to $(\leq \pi/2)$-chains here. Our use of the turn angle is more in consonance with cone production.

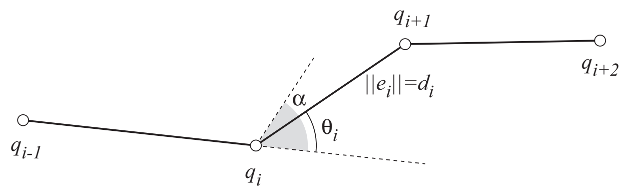

Fig. 3 Notation for a configuration Q.

A *motion* $M = \langle m_0, \ldots, m_n \rangle$ of a chain P is a list of $n+1$ continuous functions $m_i : [0, \infty] \to \mathbb{R}^3$, $i = 0, \ldots, n$, such that $M(t) = \langle m_0(t), \ldots, m_n(t) \rangle$ is a configuration of P for all $t \in [0, \infty]$. The motion is said to be *simple* if all such configurations $M(t)$ are simple. We normally assume that the motion is *finite* in the sense that, after some time T, M becomes independent of t.

2.2 Chain Production

As mentioned above, our model is that the chain is produced inside an infinite open cone C_β with apex at the origin, axis on the z axis, and half-angle (angle to the positive z-axis) β; see Fig. 2. In fact the production happens in the closure $\overline{C}_\beta$ of the cone (the cone plus its surface). Vertices and edges are produced sequentially over time inside the cone $\overline{C}_\beta$ and eventually exit through the origin. The production process maintains the invariant that at most one link, the last link produced, is (partially) inside the cone; once a link is fully outside the cone it must remain so. The last produced link must constantly touch the origin, with one half of the segment inside the cone and the other half outside the cone. The rest of the chain can move freely as long as it stays simple and never meets the cone C_β.

More precisely, at time $t_0 = 0$, the machine creates q_0 at the apex of C_β, q_1 inside $\overline{C}_\beta$, and the segment e_0 connecting them; see Fig. 4. In general, at time t_i, vertex q_i reaches the origin, and q_{i+1} and e_i are created at arbitrary locations inside the cone $\overline{C}_\beta$. The vertex q_i stays in $\overline{C}_\beta$ between times t_{i-1} and t_i, and stays outside C_β after time t_i. In total there are $n+1$ critical times satisfying $0 = t_0 < t_1 < \cdots < t_n$.

Formally, a *β-production* F is a set of $n+1$ continuous functions $f_i : [t_{i-1}, \infty] \to \mathbb{R}^3$, $i = 0, \ldots, n$, such that, for all $t \in [t_{j-1}, t_j]$, $f_j(t) \in \overline{C}_\beta$, $F(t) = \langle f_0(t), \ldots, f_j(t) \rangle$ is a simple configuration of $P[0, j]$, the segment e_{j-1} is incident to the origin, and no segment e_i intersects C_β, $i < j-1$. A configuration Q is *β-producible* if there exists a β-production F with $F(\infty) = Q$. We say that a configuration is $(\geq \alpha)$-producible if it is β-producible for some $\beta \geq \alpha$.

One consequence of this model is that, as the last link produced exits the cone $\overline{C}_\beta$, it must enter what we call the *complementary cone* $\overline{B}_\beta$. For $\beta \leq \pi/2$ (a convex cone C_β), the complementary cone $\overline{B}_\beta$ is the mirror

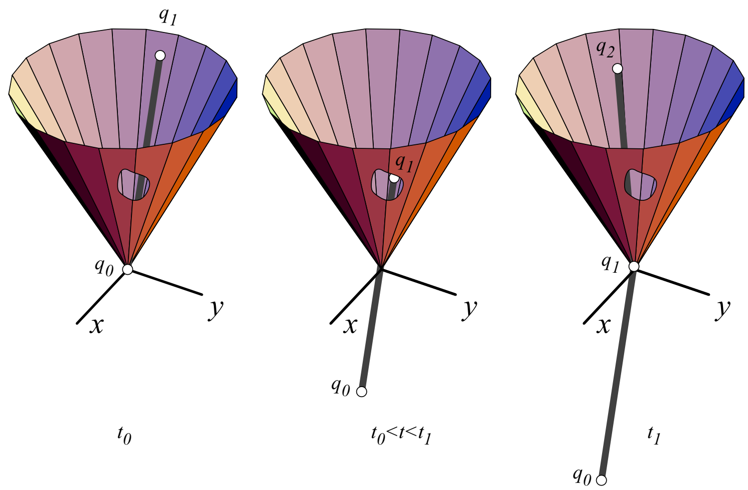

Fig. 4 Production of e_0 and e_1 during $t \in [t_0, t_1]$.

image of $\overline{C}_\beta$ with respect to the xy-plane. For $\beta \geq \pi/2$ (a reflex cone C_β), the complementary cone B_β is the region of space exterior to C_β. (This region is smaller than the mirror image of $\overline{C}_\beta$ in this case.) Fig. 5 shows an example of production when $\beta \geq \pi/2$.

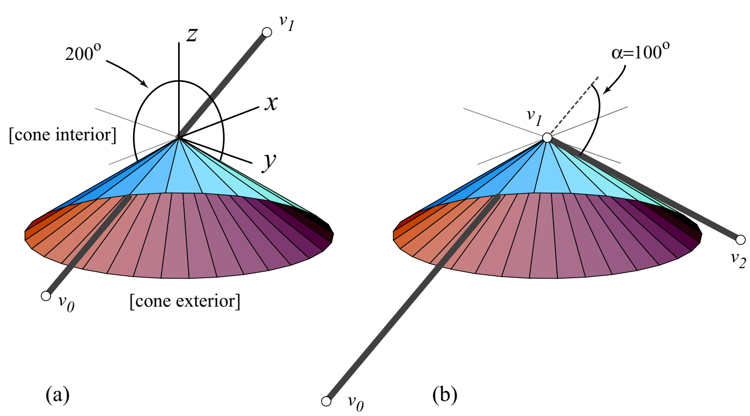

Fig. 5 Production in cone of $\beta > \pi/2$. Here $\beta = 100°$, so that the full cone angle is $200°$. The viewpoint is under the xy-plane. (a) e_0 exits to the exterior of the cone during $t \in [t_0, t_1)$. (b) e_1 is created at $t = t_1$ inside the cone, forming, in this instance, a turn angle of $100°$.

This complementary cone restricts the achievable turn angles in the producible chains:

Lemma 1 *To produce a chain whose maximum turn angle is α using a cone C_β, we must have $\alpha/2 \le \beta \le \pi - \alpha/2$.*

Proof Suppose $\theta_i = \alpha$. At time t_i, when q_{i+1} is created inside the cone, q_i is at the apex, and q_{i-1} is outside. Because we stipulate continuous motion, q_{i-1} must be inside the cone $\overline{B}_\beta$ below the xy-plane, for it must have been there throughout $t \in [t_{i-1}, t_i)$. For the same reason, q_{i+1} must be in the mirror image of $\overline{B}_\beta$ with respect to the xy-plane, because e_i is just about to enter $\overline{B}_\beta$. The cone $\overline{B}_\beta$ and its mirror image each form an angle $\min(\beta, \pi - \beta)$ with the z axis, so in order for e_{i-1} and e_i to fit those cones, $\alpha/2 \le \min(\beta, \pi - \beta)$. □

Note that arbitrarily sharp turn angles can be produced in a cone $C_{\pi/2}$, which might be viewed as a halfspace with a pinhole exit at the origin.

We will prove that there exists a simple motion between any two β-producible configurations of the same chain, and that all such configurations are flattenable. Next we define the notion of a "simple" motion.

2.3 Complexity of a Motion

There are of course many ways to define the complexity of a motion M. As a first approximation, we could assume that each dihedral angle $\delta_i^M(t)$ of the segment e_i is a piecewise-linear function of time t, and the complexity $T(M)$ of the motion M is the total number of linear pieces over all functions $\delta_i^M(t)$. That is, $T(M) = \sum_{i=1}^{n-2} T(\delta_i^M)$, where $T(\delta_i^M)$ is the number of linear pieces in the function δ_i^M. Unfortunately, this definition is not acceptable, as it restricts the range of possible motions M. The definition can be generalized to allow arbitrary functions $\delta_i^M(t)$, given some corresponding measure of complexity $T(\delta_i^M)$, with the added restriction that for every time range $t \in [r, s]$ during which $\delta_i^M(t)$ is a linear function, that time range contributes at most 1 to the complexity $T(\delta_i^M)$. For example, if $\delta_i^M(t)$ is a piecewise-polynomial function, $T(\delta_i^M)$ could be defined as the sum of the degrees of the polynomial pieces; or more generally $T(\delta_i^M(t))$ might measure the number of inflection points or monotonic pieces of $\delta_i^M(t)$.

The complexity of a production F can be defined in an analogous way, where $\delta_i^F(t)$ is defined only for the time range $t \ge t_{i+1}$. The resulting value will only account for the dihedral motions outside the cone C_β. We still need to add the complexity of the movement of point $f_{i+1}(t)$ before it exits the cone for all i, i.e., at time $t \in [t_i, t_{i+1})$. If we assume that the chain exits the cone at a constant rate, we only need to consider the vector $u^F(t) = (0, f_{i+1}(t))$ for $t \in [t_i, t_{i+1})$, described in polar coordinates by the angle $\rho^F(t)$ of $u^F(t)$ with the z-axis, and the angle $\gamma^F(t)$ of the projection of $u^F(t)$ onto the xy-plane with the x-axis. The complexity will be expressed by $T(\gamma^F)$ and $T(\rho^F)$, with the restriction that $T(\rho^F)$ be at least the number of connected components in $\{t : \rho^F(t) = 0\}$. For example, the number of pieces in a piecewise-linear function, or the sum of degrees in a piecewise-polynomial function, would qualify. We further impose on $T(\gamma^F)$ and $T(\rho^F)$ the same restriction as for $T(\delta_i^F)$. The total complexity of the production is then $T(F) = \sum_{i=1}^{n-2} T(\delta_i^F) + T(\rho^F) + T(\gamma^F)$.

3 Producible $\equiv$ Flattenable

Key to our main theorem is showing that every $(\geq \alpha)$-producible configuration of a $(\leq a)$-chain can be moved to a canonical configuration, and therefore to every other $(\geq \alpha)$-producible configuration of that chain.

3.1 Canonical Configuration

We begin by defining the canonical configuration of $(\leq \ a)$-chains, called the *α-cone canonical configuration* or *α-CCC*. To better understand the constraints of a configuration Q, consider normalizing all edge vectors $q_i q_{i+1}$ to unit vectors $u_i = (q_{i+1} - q_i)/\|q_{i+1} - q_i\|$ which lie on the unit sphere. The α-CCC is constructed to have the property that all such vectors lie along a circle of radius $\alpha/2$ on that sphere. In other words, the vectors u_i lie on the boundary of a cone with half-angle $\alpha/2$.

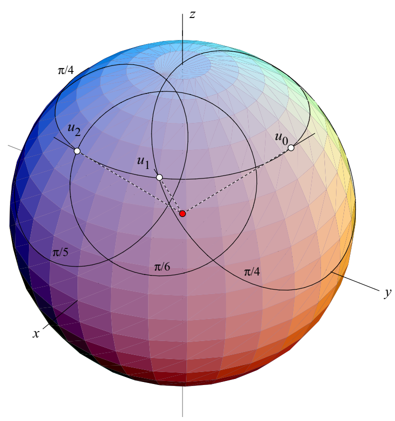

Fig. 6 u_0 lies on the cone $C_{\pi/4}$. $(\theta_1, \theta_2, \theta_3) = (\pi/4, \pi/6, \pi/5)$, respectively.

To ease the description, we use the cone $\overline{C}_{\alpha/2}$ (not C_α) to define α-CCC, but note that the cone and the chain could be rotated and translated. By convention, we place u_0 on the boundary of $\overline{C}_{\alpha/2}$ in the positive quadrant of the yz-plane. Because Q is a configuration of P, the angle between u_{i-1} and u_i is θ_i and so, on the sphere, u_i lies on the circle of radius θ_i centered at u_{i-1}. Because $\theta_i \leq \alpha$, this circle intersects the boundary of $\overline{C}_{\alpha/2}$. We set u_i to be the first intersection counterclockwise from u_{i-1} on the boundary of $\overline{C}_{\alpha/2}$ (where counterclockwise is viewed from the origin). See Fig. 6 for an example.

The position of the u_i's on the unit sphere as described above, along with the position of q_0, uniquely determine the position of the α-CCC of the chain. Because the u_i vectors all have positive z coordinates, we know that the resulting configuration is simple. See Fig. 7. We can also show that the α-CCC is completely contained in $\overline{C}_{\alpha/2}$:

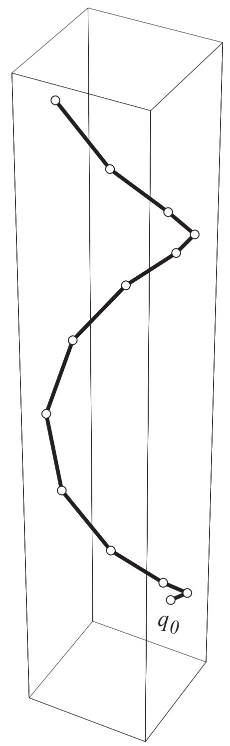

Fig. 7 A chain in its α-CCC configuration. Here $\theta_i = \pi/4$ for all i.

Lemma 2 *If all unit edge vectors u_i are contained in a cone $\overline{C}_\beta$ for some half-angle $\beta > 0$, then the configuration Q is inside $q_0 + \overline{C}_\beta$, the cone translated so its apex is at q_0. Furthermore, if $u_0 \neq u_1$, then only the first bar of the chain can touch the boundary of $q_0 + \overline{C}_\beta$.*

Proof The proof is by induction on n. The claim holds for the 1-point chain $Q[n, n]$. Assume $Q[1, n]$ is contained in a cone with apex q_1. Now q_1 is in the cone with apex q_0, so the cone with apex at q_1 is contained in the one with apex at q_0. Furthermore, the boundary of these cones intersect only if q_1 is on the boundary of $q_0 + \overline{C}_\beta$, and in that case, the intersection is contained in the line of support q_0q_1. □

In the α-CCC, u_i is always different from u_{i+1}.

3.2 Canonicalization

Next we show how to find a motion from any ($\geq \alpha$)-producible configuration of a ($\leq \alpha$)-chain to the corresponding α-CCC.

Theorem 1 *If a configuration Q of a $(\leq \alpha)$-chain P is $(\geq a)$-producible by a production F, then there is a motion M from Q to the α-CCC, with $T(M) \leq T(F) + 3n$.*

Proof Suppose that Q is β-producible for $\beta \geq \alpha$, and that F is a β-production with $F(\infty) = Q$. By scaling time appropriately, we can arrange that $t_i = i$, and the configuration freezes at time $n+1$, i.e., $F(t) = F(n+1)$ for $t > n+1$.

We construct a motion M from Q to the α-CCC, constructed inside $\overline{C}_\beta$. A key idea in our construction is to play the production movements backwards. More precisely, for all $i = 0, \ldots, n$, we define $m_i(t) = f_i(n+1-t)$ for the (reverse) time interval $t \in [0, n+2-i]$. (Beyond reverse time $n+2-i$, the original production time is less than $n+1-(n+2-i) = i-1$ and thus f_i is no longer defined.) To complete the construction, we just have to define $m_i(t)$ for $t > n+2-i$, that is, the motion of the part of the chain that has already re-entered the cone $\overline{C}_\beta$.

During the time interval $(n-i, n+1-i)$, the edge e_i is entering the cone $\overline{C}_\beta$ through the origin, $P[0,i]$ is outside C_β, and $P[i+1,n]$ is inside C_β. We maintain the invariant that $P[i,n]$ is in α-CCC, contained in a cone $\overline{C}_{\alpha/2}$ translated and rotated to some position $\overline{C}'_{\alpha/2}$. See Figure 8. So the dihedral

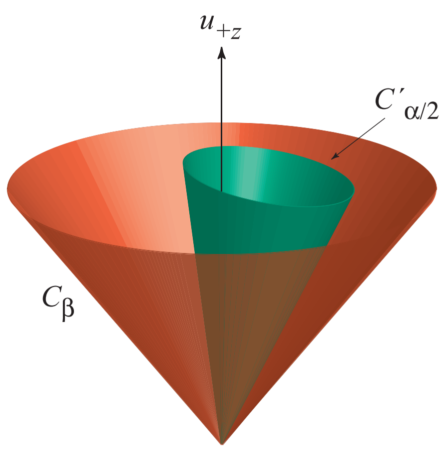

Fig. 8 Cone $C_{\alpha/2}$ is nested inside C_β. The diameter of the former is no more than the radius of the latter.

angle of e_j does not change for $j > i$, i.e., $P[i+1,n]$ is held rigid. Because $P[0,i]$ moves freely outside of C_β according to the reversed movements of the β-production, we can only control the dihedral angle of e_i in order to maintain that $\overline{C}'_{\alpha/2}$ (and so $P[i+1,n]$) stays inside $\overline{C}_\beta$.

Again, consider the vectors u_j. The invariant means that all u_j, $j = i, \ldots, n-1$, touch the boundary of some circle σ of radius $\alpha/2$ on the unit sphere centered on the apex of the cone, and σ must be inside $\overline{C}_\beta$. For any position u_i, we place σ so that its center is on the great arc between u_i and u_{+z}, where u_{+z} is the unit vector along the the z-axis. This implies that u_i is the farthest point from u_{+z} on σ and since, by the production constraints, u_i is in $\overline{C}_\beta$, σ is in $\overline{C}_\beta$ as well and the invariant is satisfied. As long as $u_i \neq u_{+z}$, this position of σ is unique and the resulting motion is continuous because the production is continuous. When $u_i = u_{+z}$, a discontinuity might be introduced, but these discontinuities can easily be removed by stretching the moment of time at which a discontinuity occurs and filling in a continuous motion between the two desired states.

At time $t = n + 1 - i$, vertex i enters $\overline{C}_\beta$ and the invariant needs to be restored for the next phase. At that time, the vector u_{i-1} lies in $\overline{C}_\beta$, and u_i is on a circle τ of radius θ_i centered at u_{i-1}. Let σ' be the desired new position for σ, that is, the circle whose radius is $\alpha/2$, and whose center is on the great arc between u_{i-1} and u_{+z}. We know that σ' and τ intersect and all intersections are inside $\overline{C}_\beta$ because σ' is in $\overline{C}_\beta$. See Figure 9(a). We first move u_i along τ to the first intersection between σ' and τ counterclockwise from u_{i-1} on σ' (Figure 9(b)) by changing the dihedral angle of e_{i-1}, and simultaneously moving σ accordingly as described above by changing the dihedral angle of e_i. This can be done while maintaining the invariant because the intersection of τ and $\overline{C}_\beta$ is connected. We then rotate σ about u_i to the position σ' (Figure 9(c)) by changing the dihedral angle of e_i. This motion can be done while maintaining the invariant because the set of dihedral angles of e_i for which σ is in $\overline{C}_\beta$ is connected.

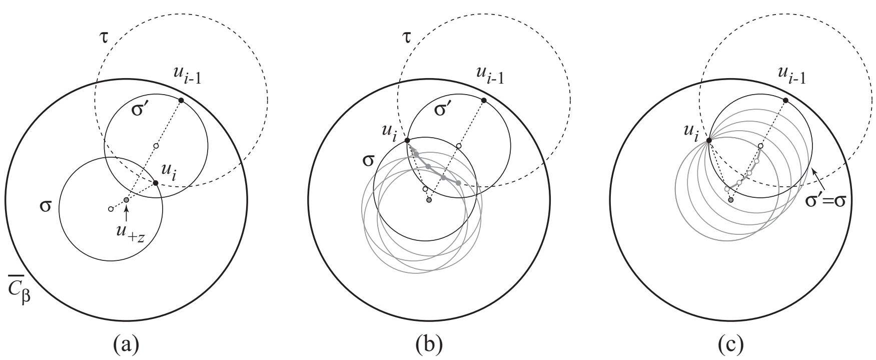

Fig. 9 Restoring the invariant. View looking down u_{+z}. (a) σ and σ' are both radius $\alpha/2$, determined by $\overline{C}_{\alpha/2}$, which moves inside $\overline{C}_\beta$, centered on u_{+z}. τ is of radius θ_i. (b) u_i walks to the ccw point of $\sigma' \cap \tau$. (c) σ is rotated about u_i. Here $\alpha/2 = 30^\circ < \theta_i = 50^\circ < \alpha = 60^\circ < \beta = 70^\circ$.

The complexity of all dihedral motions outside of C_β is $\sum_{i=1}^{n-2} T(\delta_i^F)$. The dihedral motions of e_i during times $t \in (n - i, n + 1 - i)$ mirror exactly $\gamma^F(n+1-t)$, except at discontinuities, which correspond to times for which $u_i = u_{+z}$, which is exactly when $\rho^F(n + 1 - t) = 0$, so the total complexity of these dihedral motions is bounded by $T(\rho^F) + T(\gamma^F)$. Finally, whenever a vertex attains the apex of the cone, we perform three dihedral rotations (linear functions of time) to restore the invariant. Summing it all, we obtain $T(M) \leq \sum_{i=1}^{n-2} T(\delta_i^F) + T(\rho^F) + T(\gamma^F) + 3n = T(F) + 3n$. □

Corollary 1 *For any two simple ($\geq \alpha$)-producible configurations Q_1 and Q_2 of a common ($\leq \alpha$)-chain, with respective productions F_1 and F_2, there is a simple motion M from Q_1 to Q_2—that is, $M(0) = Q_1$ and $M(\infty) = Q_2$—for which $T(M) \leq T(F_1) + T(F_2) + 6n$.*

Proof Because Q_1 and Q_2 are ($\geq \alpha$)-producible, the previous theorem gives us two motions M_1 and M_2 with $M_1(0) = Q_1$, $M_1(\infty) = \alpha$-CCC, $M_2(0) = Q_2$, and $M_2(\infty) = \alpha$-CCC. By rescaling time, we can arrange that $M_1(t) = M_2(t) = \alpha$-CCC for t beyond some time T. Then define $M(t) = M_1(t)$ for $0 \leq t \leq T$, $M(t) = M_2(2T - t)$ for $T < t \leq 2T$, and $M(t) = Q_2$ for $t > 2T$. □

Lemma 3 *An α-CCC of a ($\leq \alpha$)-chain is β-producible for any $\alpha/2 \leq \beta \leq \pi - \alpha/2$. The complexity of the production is at most $2n - 1$.*

Proof Let Q be a α-CCC positioned in $\overline{C}_{\alpha/2}$ with q_0 at the origin. Let $q(t)$ be the point at distance t from q_0 along Q. The position $F(t)$ of the produced portion of the α-CCC at time t is Q translated so that $q(t)$ is at the origin and deleting all the edges of Q completely inside $C_{\alpha/2}$. By Lemma 2, all edges of $F(t)$ except for the edge containing the origin are contained in the cone $B_{\alpha/2}$. F is thus a valid β-production for any $\alpha/2 \leq \beta \leq \pi - \alpha/2$. The β-production doesn't use any dihedral rotation so $T(\delta_i^F) \leq 1$, $\rho^F(t) = \alpha/2$ for all t so $T(\rho^F) \leq 1$, and γ^F is constant for every edge, so $T(\gamma^F) \leq n$ □

Corollary 2 *If a configuration Q of a ($\leq \alpha$)-chain has a β-production F for some $\beta \geq \alpha$, then it has a β'-production F' for all $\alpha/2 \leq \beta' \leq \pi - \alpha/2$ and $T(F') \leq T(F) + 5n + 1$.*

Proof Using Theorem 1, let M be the motion from Q to an α-CCC, and let M' be the reverse motion from the α-CCC to Q. Let R be the sum of the edge lengths of the chain. The production F' first produces a α-CCC in $B_{\alpha/2}$ using Lemma 3. The α-CCC is then translated by a distance $R/\sin\alpha/2$ in the negative direction along the z axis. At this point, the sphere centered at q_n and of radius R doesn't intersect the outside of $B_{\alpha/2}$. Keeping q_n fixed, we perform the motion M' to obtain configuration Q. □

3.3 Connection to Flat States

Finally, we relate flat configurations to productions and prove our main result that flattenability is equivalent to producibility.

Lemma 4 *All flat configurations of a ($\leq \alpha$)-chain have a β-production F for any β satisfying $\alpha \leq \beta \leq \pi/2$. Furthermore, $T(F) \leq n$.*

Proof Assume the configuration is in the xy-plane. Any such flat configuration can be created using the following process. First, draw e_0 in the xy-plane. Then, for all consecutive edges e_i, create e_i in the vertical plane through e_{i-1} at angle θ_{i-1} with the xy-plane, then rotate it to the desired position in the xy-plane by moving the dihedral angle of e_{i-1}. During the creation and motion of e_i, it is possible to enclose it in some continuously moving cone C of half-angle β whose interior never intersects the xy-plane: at the creation of e_i, C is tangent to the xy plane on the support line of e_{i-1} and with its apex at p_i, and thus contains e_i. During the rotation of e_i, e_i will eventually touch the boundary of C. We then move C along with e_i so that both e_i and the xy-plane are tangent to C. When e_i reaches the xy plane, we translate C along e_i until its apex is p_{i+1}. Viewing the construction relative to C and placing C on C_β gives the desired β-production. □

Corollary 3 ($\leq \pi/2$)*-chains are flat-state connected. The motion between any two flat configurations uses at most* $8n$ *dihedral motions.*

Proof Consider two flat configurations Q and Q' of a ($\leq \pi/2$)-chain. By Lemma 4, Q and Q' are both ($\pi/2$)-producible, and so by Corollary 1, there exists a motion M such that $M(0) = Q$ and $M(+\infty) = Q'$. □

Corollary 4 *All* α*-producible configurations of* ($\leq \alpha$)*-chains are flattenable, provided* $\alpha \leq \pi/2$*. For a production* F*, the flattening motion* M *has complexity* $T(M) \leq T(F) + 7n$*.*

Proof Consider an α-producible configuration Q of an ($\leq \alpha$)-chain P. Because $\alpha \leq \pi/2$, the chain P also has a flat configuration Q' [ADD$^+$02]. By Lemma 4, Q' is producible, and so by Corollary 1, there exists a motion M such that $M(0) = Q$ and $M(+\infty) = Q'$. The bound on $T(M)$ is by composition of the bounds in Lemma 4 and Corollary 1. □

We note that the restriction in our results to $\alpha \leq \pi/2$ accords with the generally obtuse (about 110°) protein bond angles, which correspond to turn angles α of about 70°.

4 A More Powerful Machine

We now show that our result does not hold without the assumption $\alpha \leq \beta$, under a somewhat stronger model of production that also breaks Lemma 1 that $\alpha \leq 2\beta$.

The stronger model of production separates the creation of the next vertex v_{i+1} from the moment that the previous vertex v_i reaches the origin. Specifically, we suppose that v_{i+1} is not created at t_i, but rather imagine the time instant t_i to be stretched into a positive-length interval $[t_i, t'_i]$, allowing time for $v_i v_{i-1}$ to rotate exterior to the cone prior to the creation of v_{i+1} (at time t'_i). This flexibility removes the connection in Lemma 1 between the half-angle β of the cone and the turn angles α produced, permitting chains of large turn angle from any cone. Indeed, the sequence of motions depicted in Fig. 10 exploits this large-angle freedom to emit a 4-link fixed-angle chain that is locked.

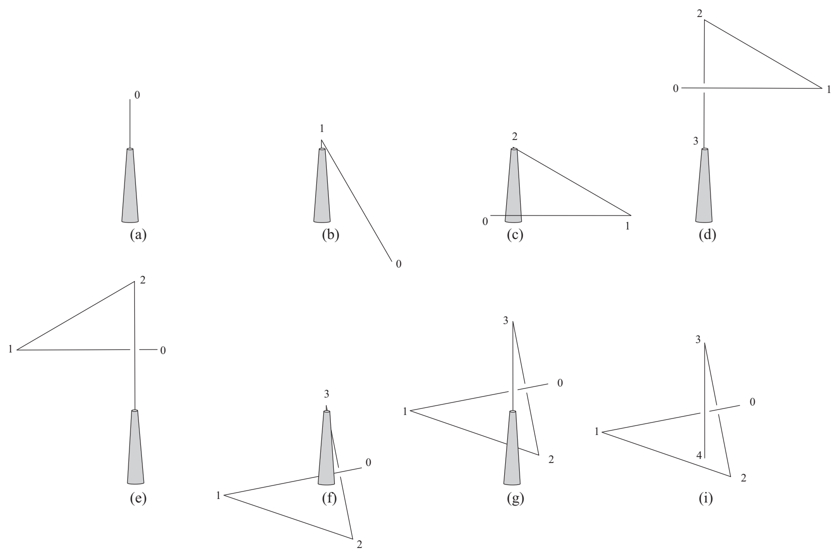

Fig. 10 Production of a locked chain under a model that permits large turning angles to be created. For clarity, the cone is reflected to aim upward. (a) $e_0 = (q_0, q_1)$ emerges; (b) turn at q_1; (c) turn at q_2 and dihedral motion at q_1 places e_1 in front of cone; (d) e_2 nearly fully produced; (e) chain spun about e_2 (or viewpoint changed); (f) rotation at q_3 away from viewer places chain behind cone; (g) e_3 emerges; (i) final locked chain shown loose; the turn angle θ_3 at q_3 can be made arbitrarily close to π.

5 Random Chains

This section proves that the producible/flattenable configurations are a vanishingly small subset of all possible configurations of a chain, for almost any chain. Essentially, the results below say that, if there is one configuration of one chain in a class that is unflattenable, then a randomly chosen configuration of a randomly chosen chain from that class is unflattenable with probability approaching 1 geometrically as the number of links in the chain grows. Furthermore, this result holds for any "reasonable" probability distribution on chains and their configurations.

To define probability distributions, it is useful to embed chains and their configurations into Euclidean space. A chain $P = \langle \theta_1, \ldots, \theta_{n-1}; d_0, \ldots, d_{n-1} \rangle \in [0, \pi/2]^{n-1} \times [0, \infty)^n$ is specified by its turn angles θ_i and edge lengths d_i. A configuration $Q = \langle \delta_1, \ldots, \delta_{n-2} \rangle \in [0, 2\pi)^{n-2}$

of P is specified by its dihedral angles. We also need to be precise about our use of the term "unflattenable" for chains vs. configurations. A simple configuration Q is *unflattenable* or simply *locked* if it cannot reach a flat configuration; a chain P is *lockable* if it has a locked configuration.

We consider the following general model of random chains of size n. Call a probability distribution *regular* if it has positive probability on any positive-measure subset of some open set called the *domain*, and has zero probability density outside that domain.[3] For Euclidean d-space $\mathbb{R}^d$, a probability distribution is regular if it has positive probability on any positive-radius ball inside the domain. Uniform distributions are always regular.

For chains of k links, we emphasize the regular probability distribution $\mathcal{P}_k^{\Theta,\mathcal{D}}$ obtained by drawing each turn angle θ_i independently from a regular distribution Θ, and drawing each edge length d_i independently from a regular distribution $\mathcal{D}$. Similarly, for not-necessarily-simple configurations of a fixed chain P, we emphasize the regular probability distribution obtained by drawing each dihedral angle δ_i independently from a regular distribution Δ. We can modify this probability distribution to have a domain of all simple configurations of P instead of all configurations of P, by zeroing out the probability density of nonsimple configurations, and rescaling so that the total probability is 1. The resulting distribution is denoted $\mathcal{Q}^{P,\Delta}$, and it is regular because of the following well-known property:

Lemma 5 *The subspace of simple configurations of a chain P is open.*

Proof Consider the space $[0, 2\pi)^{k-2}$ of all configurations of P. The simplicity of a configuration Q of P can be expressed by the $O(k^2)$ constraints that no two nonadjacent segments intersect. These (semi-algebraic) constraints are all of the form $g(Q) < 0$ where $g(Q) = g(\delta_1, \ldots, \delta_{k-2})$ is a multinomial of a constant number of terms in $\sin(\delta_i)$ and $\cos(\delta_i)$. Each constraint defines an open set in the configuration space. The conjunction of the constraints corresponds to the intersection of these finitely many sets, which is open. □

First we show that individual locked examples immediately lead to positive probabilities of being locked. The next lemma establishes this property for configurations of chains, and the following lemma establishes it for chains.

Lemma 6 *For any regular probability distribution $\mathcal{Q}$ on simple configurations of a lockable chain P, if there is a locked simple configuration in the domain of $\mathcal{Q}$, then the probability of a random simple configuration Q of P being locked is at least a constant $c > 0$.*

Proof Let Q' be a locked simple configuration in the domain of $\mathcal{Q}$. Let C be the component of the space of simple configurations containing Q', and let D be the intersection of C and the domain of $\mathcal{Q}$. Because C is open and

[3] A closely related but more specific notion of regular probability distributions in 1D was introduced by Willard [Wil85] in his extensions to interpolation search.

thus D is open, there exists a constant $\varepsilon > 0$ such that the ball B of radius ε centered at Q' is contained in D, and all $Q'' \in B$ are locked as well. Choose c to be the probability of choosing a configuration in B, which is positive by regularity. $\square$

Lemma 7 *For any regular probability distribution $\mathcal{P}$ on chains, if there is a lockable chain in the domain of $\mathcal{P}$, then the probability of a random chain P being lockable is at least a constant $\rho > 0$.*

Proof Consider the space of all chains and configurations of those chains, $\mathcal{C} = [0, \pi/2]^{n-1} \times [0, \infty)^n \times [0, 2\pi)^{n-2}$. As described in Lemma 5, the constraint that a particular configuration is locked can be phrased as a set of open semi-algebraic constraints, except now the constraints depend on all $3n - 3$ variables (not just the dihedral angles). Intersecting all these open semi-algebraic sets results in a subspace $\mathcal{L} \subset \mathcal{C}$ of all locked configurations of all chains. Projecting this open set down to $\mathcal{L}' \subseteq [0, \pi/2]^{n-1} \times [0, \infty)^n$ by dropping the dihedral angles results in another open semi-algebraic set, because open semi-algebraic sets are closed under projection.

Now let P' be a lockable chain in the domain of $\mathcal{P}$, let C be the component of $\mathcal{L}'$ containing P', and let D be the intersection of C and the domain of $\mathcal{P}$. Because C and thus D is open, there is a constant $\varepsilon > 0$ such that the ball B of radius ε centered at P' is contained in D, and all $P'' \in B$ are lockable. Choose ρ to be the probability of choosing a chain in B, which is positive by regularity. $\square$

Next we show that these positive-probability examples of being locked lead to increasing high probabilities of being locked as we consider larger chains.

Theorem 2 *Let P_n be a random chain drawn from the regular distribution $\mathcal{P}_n^{\Theta,\mathcal{D}}$. If there is a lockable chain in the domain of $\mathcal{P}_n^{\Theta,\mathcal{D}}$ for at least one value of n, then*

$$\lim_{n\to\infty} Pr[P_n \textit{ is lockable}] = 1.$$

Furthermore, if Q_n is a random simple configuration drawn from the regular distribution $\mathcal{Q}^{P_n,\Delta}$, then

$$\lim_{n\to\infty} Pr[Q_n \textit{ is flattenable}] = \lim_{n\to\infty} Pr[Q_n \textit{ is producible}] = 0.$$

Both limits converge geometrically.

Proof Suppose there is a lockable chain of k links. By Lemma 7, $\Pr[P_k \text{ is lockable}] > \rho > 0$. Break P_n into $\lfloor n/k \rfloor$ subchains of length k. Each of these subchains is chosen independently from $\mathcal{P}_k^{\Theta,\mathcal{D}}$ and is not lockable with probability $< 1 - \rho$. Now P_n is lockable (in particular) if any of the subchains are lockable, so the probability that P_n is not lockable is $< (1 - \rho)^{\lfloor n/k \rfloor}$ which approaches 0 geometrically as n grows. Likewise, by Lemma 6, the probability that Q_k is locked is $> c\rho$ for some constant $0 < c < 1$, and so the probability that Q_n is flattenable is $< (1 - c\rho)^{\lfloor n/k \rfloor}$ which approaches 0 geometrically as n grows. $\square$

Thus, producible configurations of chains become rare as soon as one chain in the domain of the distribution is lockable. The locked "knitting needles" example of [CJ98,BDD+01] can be built with chains satisfying $\alpha \le \pi/2$ by replacing the acute-angled universal joints with obtuse, fixed-angled chains of very short links. Thus for any regular distribution including such examples in its domain, we know that configurations of $(\le \alpha)$-chains are rarely producible for the case we have considered, $\alpha \le \pi/2$. We do not know of any nontrivial regular probability distribution $\mathcal{P}_n^{\Theta,\mathcal{D}}$ whose domain has no lockable chains. In particular, for equilateral (all edge-lengths equal) fixed-angle chains, it is not known whether angle restrictions can prevent the existence of locked configurations. As protein backbones are nearly equilateral, it is of particular interest to answer this question.

Future directions for research include resolving the locked question just mentioned, incorporating the short side-chains that jut from the protein backbone, and more realistically modeling the ribosome structure.

Acknowledgements

Much of this work was completed at the *Workshop on Geometric Aspects of Molecular Reconfiguration* organized by Godfried Toussaint at the Bellairs Research Institute of McGill University in Barbados, February 2002. We appreciate the helpful discussions with the other participants: Greg Aloupis, Prosenjit Bose, David Bremner, Vida Dujmović, Herbert Edelsbrunner, Jeff Erickson, Ferran Hurtado, Henk Meijer, Pat Morin, Mark Overmars, Suneeta Ramaswami, Ileana Streinu, Godfried Toussaint, and especially Yusu Wang.

References

ADD+02. G. Aloupis, E. Demaine, V. Dujmović, J. Erickson, S. Langerman, H. Meijer, I. Streinu, J. O'Rourke, M. Overmars, M. Soss, and G. Toussaint. Flat-state connectivity of linkages under dihedral motions. In *Proc. 13th Annu. Internat. Sympos. Alg. Comput.*, volume 2518 of *Lecture Notes in Comput. Sci.*, pages 369–380. Springer, 2002.

ADM+02. Greg Aloupis, Erik D. Demaine, Henk Meijer, Joseph O'Rourke, Ileana Streinu, and Godfried Toussaint. Flat-state connectedness of fixed-angle chains: Special acute chains. In *Proc. 14th Canad. Conf. Comp. Geom.*, pages 27–30, 2002.

BDD+01. T. Biedl, E. Demaine, M. Demaine, S. Lazard, A. Lubiw, J. O'Rourke, M. Overmars, S. Robbins, I. Streinu, G. Toussaint, and S. Whitesides. Locked and unlocked polygonal chains in 3D. *Discrete Comput. Geom.*, 26(3):269–282, 2001.

CJ98. J. Cantarella and H. Johnston. Nontrivial embeddings of polygonal intervals and unknots in 3-space. *J. Knot Theory Ramifications*, 7(8):1027–1039, 1998.

NHB+00. P. Nissen, J. Hansen, N. Ban, B. Moore, and T. A. Steitz. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289:920–930, August 2000.

ST00. M. Soss and G. T. Toussaint. Geometric and computational aspects of polymer reconfiguration. *J. Math. Chemistry*, 27(4):303–318, 2000.

Wil85. D. E. Willard. Searching unindexed and nonuniformly generated files in $\log\log N$ time. *SIAM J. Comput.*, 14:1013–1029, 1985.