Concatenative nonmonotonicity of optimal protein folding

Bjørn Kjos-Hanssen

University of Hawai‘i at Mānoa, Honolulu HI 96822, USA
https://math.hawaii.edu/wordpress/bjoern/ bjoernkh@hawaii.edu

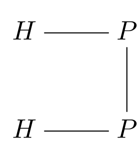
Abstract. The hydrophobic-polar (HP) protein folding model was introduced by Ken A. Dill in 1985. In this model, a binary string like 0110 is interpreted as a polymer, a sequence HPPH of amino acids of two types. The string is embedded in an ambient space, and this embedding is called a fold. Some folds are better than others and this is quantified by the score of the fold.
We prove that the optimal score function is not monotonic under concatenation for the 2D rectangular lattice, the 3D rectangular lattice, the hexagonal lattice, and the triangular lattice. In other words, the concatenation of two polymers may have strictly lower minimum energy than either polymer.

Keywords: Hydrophobic-polar protein folding model, monotonicity, concatenation

1 Introduction

Protein folding prediction has long been considered an intractable problem. In order to make this precise, in 1985 Ken A. Dill [6] introduced a mathematical model, the hydrophobic-polar (HP) protein folding model. It was no real surprise when the NP-completeness of its optimality problem was demonstrated in both the two- and three-dimensional settings [4,3]. Google's AlphaFold [10] showed that in practical terms protein folding is more feasible than the NP-completeness would suggest. However, the HP model remains mathematically fascinating.

Its basic setup is that a protein is modeled as a word w from the alphabet $\{H, P\}$. A fold is a self-avoiding walk in a lattice graph labeled by consecutive entries of w. When two non-consecutive occurrences of H are connected by an edge in the lattice, a point is achieved:



An optimal fold is one that achieves the maximum number of points. (In this article we identify $H = 0$ and $P = 1$.)

We can view this as an example of parametric optimization. For a word $\theta \in \Theta \in \{0,1\}^*$, and folds x, the *optimal value* function J is given by

$$J(\theta) = \max_x f(x, \theta),$$

where the *fitness function* or *objective function* f gives the score of θ under the fold x. If we write $f = -E$ where E stands for *energy*, the task is to minimize energy.

The folds x may equivalently be viewed as either sequences of consecutive locations in the lattice, or sequences of "moves" describing how to get the next locations from the previous.

Combinatorial embeddings of graphs. A combinatorial embedding of a graph is a pair of permutations (v, e) which act on a set B of $2|E(G)|$ darts. The permutation e is an involution, and its orbits correspond to edges in the graph. Similarly, the orbits of v correspond to the vertices of the graph, and those of $f = ve$ correspond to faces of the embedded graph.[1] A move is a number k such that the next dart is $v^k e$ of the current dart. However, traditionally in the literature on protein folding one considers *actual* embeddings into specific lattices like $\mathbb{Z}^2$ rather than combinatorial embeddings; hence we shall say no more on the topic.

Automatic complexity. There is an interesting similarity between the HP model and automatic complexity [11]. We can view a point scored as a negative state. When a point is scored, the polymer is returning to a previous location, analogously to how an automaton returns to a state. Minimizing the number of states corresponds to maximizing the number of returns to previous states, which is (except for the division into hydrophobic and polar amino acids) analogous to minimizing the energy which is the negative of the score.

Overview of the paper. In Section 2, we demonstrate that the optimal score in the hydrophobic-polar protein folding model is non-monotone under concatenation, in several common variants of the model: 2D, 3D, triangular, and hexagonal grids.

For 2D rectangular and hexagonal we can prove this as a corollary of constructing an optimal fold in terms of the number of hydrophobic amino acids $Z(w)$.

For 2D rectangular, 3D rectangular, and triangular lattice we can prove it by a direct construction using isoperimetric inequalities.

See Table 1 for an overview of these results.

In Section 3 we show that this nonmonotonicity is prevalent enough to obtain an NP-hardness result. Finally, in Section 4 we obtain optimal closed folds with knots and links in the 3D HP model.

[1] https://doc.sagemath.org/html/en/reference/graphs/sage/graphs/genus.html

Lattice	Proof using Z-optimality	Isoperimetric fact	Proof using that fact
triangular	not known	Theorem 9	Theorem 10
hexagonal	Theorem 6	not possible	
2D rectangular	Theorem 1	Theorem 13	Theorem 14
3D rectangular	not known	Theorem 12	Theorem 15

Table 1: Nonmonotonicity results for optimal protein folding in HP models.

2 Nonmonotonicity

Definition 1. *Let $J_{\mathrm{rect}}(x), J_{\mathrm{cube}}(x), J_{\mathrm{tri}}(x), J_{\mathrm{hex}}(x)$ denote the maximum number of points achievable for a word x in the 2D and 3D rectangular, triangular, and hexagonal lattices, respectively.*

In a March 5, 2023, email message [12], Jack Stecher conjectured that the HP folding problem is "weakly monotone", i.e., that whenever we add a prefix or suffix to a given word, our optimal score should never decrease.

Conjecture 1 (Stecher's cul-de-sac *conjecture).* Let x, y be binary words. Then $J_{\mathrm{rect}}(x) \leq J_{\mathrm{rect}}(xy)$.

Clearly the reverse x^R of a word x satisfies $J(x) = J(x^R)$. Therefore, we could equivalently conjecture that $J_{\mathrm{rect}}(y) \leq J_{\mathrm{rect}}(xy)$. To refute the conjecture we make use of Theorem 1, which is also of independent interest.

Theorem 1. *There are infinitely many words w with $J_{\mathrm{rect}}(w) = Z(w)+1$, where $Z(w)$ is the number of zeros in w.*

Proof. There are examples of each length of the form $26 + 8k$, $k \geq 0$:

$$(011)^3 1^{10} (0011)^k 0110 (1100)^k 110 \tag{1}$$

A suitable fold is defined by the $25 + 8k$ moves

$$urdrdldrru^4 l^4 dd(luld)^k ldr(drur)^k dru,$$

where u is "up", l is "left", d is "down", and r is "right". The case $k = 3$ is shown in Figure 1. See Figure 2 for a witness for the hexagonal lattice.

Theorem 2 (Agarwala et al. [1, Lemma 2.1]). *For all w, $J_{\mathrm{rect}}(1w1) \leq Z(w)$.*

Proof. Each zero in $x = 1w1$ is internal, i.e., it is neither the first nor the last bit of x. Therefore, among the **four** "heavenly" directions from the location of the zero, two are occupied by the previous and the next amino acids, and the other **two** could each contribute to a point. As each edge is shared by two vertices, by the Handshaking Lemma each zero contributes on average at most **one** point.

When the word may start or end with a zero the bound becomes:

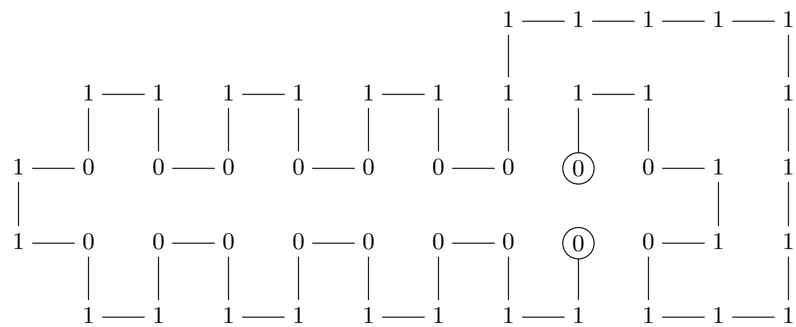


Fig. 1: A 2D rectangular fold achieving the maximum score in terms of the number of zeros. While this result appears to be new, a special case with $Z = 6$ was found by Matthew Gilzinger [7, Figure 13].

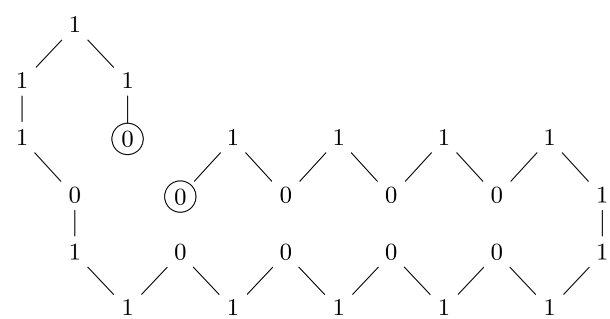


Fig. 2: A hexagonal lattice fold achieving the maximum score $Z(w)/2 + 1 = 6$ in terms of the number of zeros $Z(w) = 10$ in the folded word.

Theorem 3 ([2, Fact 1]). *For all w, $J_{\text{rect}}(w) \leq Z(w) + 1$.*

Proof. There are now up to **three** available heavenly directions where a point may be earned at the first and last bits of w. The upper bound is therefore, with $w = w_1 \dots w_n$, each $w_i \in \{0,1\}$,

$$\frac{3 \cdot 1_{w_1=0} + (\sum_{i=2}^{n-1} 2 \cdot 1_{w_i=0}) + 3 \cdot 1_{w_n=0}}{2} \leq \left(\sum_{i=1}^{n} 1_{w_i=0} \right) + 1 = Z(w) + 1.$$

Here, the indicator symbol 1_P is 1 if P holds and 0 otherwise.

Corollary 1. *The* cul-de-sac *conjecture is false.*

Proof. Suppose otherwise. Let x be a word of the form (1). Then we have the arithmetic contradiction

$$\begin{array}{rcrr} Z(x)+1 & = & J_{\text{rect}}(x) & \text{by Theorem 1,} \\ & \leq & J_{\text{rect}}(1x1) & \text{by the } \textit{cul-de-sac} \text{ conjecture,} \\ & \leq & Z(x) & \text{by Theorem 2.} \end{array}$$

The vertices of the *hexagonal lattice* are vertices of hexagons and form a 3-regular graph.

Theorem 4. *For all w, $J_{\text{hex}}(1w1) \leq Z(w)/2$.*

Proof. Each zero in $x = 1w1$ is internal, i.e., it is neither the first nor the last bit of x. Therefore, among the **three** "heavenly" directions from the location of the zero, two are occupied by the previous and the next amino acids, and the other **one** could contribute to a point. As each edge is shared by two vertices, by the Handshaking Lemma each zero contributes on average at most **one half** of a point.

When the word may start or end with a zero the bound becomes:

Theorem 5. *For all w, $J_{\text{hex}}(w) \leq Z(w)/2 + 1$.*

Proof. There are now up to **two** available heavenly directions where a point may be earned at the first and last bits of w. The upper bound is therefore, with $w = w_1 \dots w_n$, each $w_i \in \{0,1\}$,

$$\frac{2 \cdot 1_{w_1=0} + (\sum_{i=2}^{n-1} 1 \cdot 1_{w_i=0}) + 2 \cdot 1_{w_n=0}}{2} \leq \frac{1}{2} \left(\sum_{i=1}^{n} 1_{w_i=0} \right) + 1 = Z(w)/2 + 1.$$

Theorem 6. *For each integer $k \geq 1$, the word $w_k = 01^4011(01)^k 11(10)^k$ satisfies $J_{\text{hex}}(w_k) = Z(w_k)/2 + 1$.*

Proof. Using the fold in (2) we obtain $J_{\text{hex}}(w) = Z(w)/2 + 1$, which is the best possible by Theorem 4.

Corollary 2. *There are infinitely many counterexamples to the* cul-de-sac *conjecture for the hexagonal lattice.*

The proof of Corollary 2 has the same steps as the proof of Corollary 1.

Theorem 7. *Let x be a binary word and $J \in \{J_{\text{rect}}, J_{\text{tri}}, J_{\text{hex}}\}$. Then $J(x1) \leq J(x)$ and $J(1x) \leq J(x)$.*

Proof. If F is a fold achieving maximum score for $x1$ then we obtain a fold of x with the same score simply by removing the final (respectively, initial) 1 from the fold.

Theorem 8 provides one way of going from one counterexample to infinitely many, in the case of the *cul-de-sac* conjecture. In general, such an improvement is of interest because it necessarily requires insight beyond brute force computation.

Theorem 8. *For $J \in \{J_{\text{rect}}, J_{\text{tri}}, J_{\text{hex}}\}$, if $J(x1) < J(x)$ is witnessed by a fold F of x which can be extended to a fold of $1^m x$ for some $m \in \mathbb{N}$ then $J(1^m x1) < J(1^m x)$ as well.*

Proof. Using Theorem 7 and our assumptions we have

$$J(1^m x1) \leq J(x1) < J(x) = J(1^m x).$$

Theorem 9 records the fact that if $n = 1 + 3r(r+1)$ then $D_n = B_r$ (the closed ball with radius r, also known as the hexagonal daisy D_n) is a solution to the *edge-isoperimetric problem* for cardinality n. This means that the minimum number of edges connect vertices in the set B_r to vertices outside the set, among all sets of cardinality n.

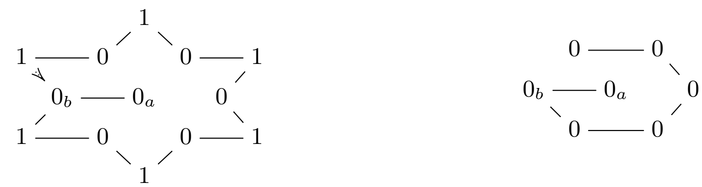
Theorem 9 ([8, Corollary 7.1]; see also [5]). *For $n = 1 + 3r + 3r^2$, the hexagonal daisy D_n is the unique minimizer of the edge-isoperimetric problem for the triangular lattice.*

Given words η, θ, we define $I_\eta(\theta) = \text{occ}(\theta, \eta)$ to be the number of occurrences of η in θ. For example, $I_{00}(001000) = 3$. The statistic $I_{00}(\theta)$ represents the number of internal "hydrophobic" connections of θ.

Theorem 10. *There are infinitely many counterexamples to the* cul-de-sac *conjecture for the triangular lattice.*

Proof. We shall exhibit for each $n \in \mathbb{N}$ a pair (x, w) of a fold x and a word w such that

- w contains no subword 0^{2n+1}, and
- for each word v with $Z(v) = Z(w)$, if $I_{00}(1v1) + f(y, 1v1) \geq I_{00}(w) + f(x, w)$ then v contains a subword 0^{2n+1}.



(a) Pair (x, ζ) with $I_{00}(\zeta) = 1, f(x, \zeta)) = 11$, $\zeta = 0(01)^6$.

(b) Pair (y, η) with $I_{00}(\eta) = 6$, $f(y, \eta) = 6$, $\eta = 0^7$.

Fig. 3: Illustration of the Induced Edge Problem solution for the triangular lattice.

(We apply this with $v = w$.) The fold x will be such that the zeros of w are organized in a metric ball, which forces y to do the same for $1v1$.

Indeed, by Theorem 9, for any words η, θ with $Z(\eta) = Z(\theta)$ and where $I_{00}(\zeta) \geq I_{00}(\eta)$, and any folds x, y, if x has the zeros of η in a metric ball then $f(x, \eta) \geq f(y, \zeta)$.

This is because in general, if the zeros of a word ζ are organized in a metric ball by a fold x then

$$I_{00}(\zeta) + f(x, \zeta) = \max_{\eta : Z(\eta) = Z(\zeta)} I_{00}(\eta) + f(y, \eta).$$

For example, the pairs $(x, \zeta), (y, \eta)$ in Figure 3 both realize the maximum above.[2]

Therefore, in any fold y of $1v1$ with $f(y, 1v1) = J(1v1)$, letting y_0 be the induced fold of v, we have $f(y_0, v) \geq f(y, 1v1) \geq f(x, w)$ and so the zeros of v must also be organized in a metric ball in y_0. Moreover, in order for $1v1$ to achieve enough points it must be that $Z(v) = Z(w)$. Write $v = u0w$ where the 0 between u and w is located in the center of the metric ball according to the fold y. Then the path from the initial 1 of $1v1$ to the final 1 of $1v1$ via the center 0 of the ball must contain a factor that enters the metric ball, proceeds to its center, and exits the metric ball. Thus, we must have $u = u'0^n$ and $w = 0^n w'$ for some words u', w'. We conclude that v contains a factor of the form 0^{2n+1}.

We create some folds of 0^* in the shape of a metric ball, and add *whiskers*, replacing some edges $0 \to 0$ on the boundary by $0 \to 1^* \to 0$, in order to prevent long sequences of 0s. For example, the fold on the left in (3) has whiskers.

We start from a closed loop configuration and let the arrow leaving the 0 indicated by 0_a be moved from a "symmetric" position indicated by ... to 0_b. See Figure 4, Figure 5 for the pairs $P_n = (x_n, w_n)$, $n \leq 5$.

[2] This sum of I_{00} and f is discussed in the second paragraph of Section 2 by Berger and Leighton, and on page 3 by Harper where he phrases it as: *the Edge-Isoperimetric Problem and the Induced Edge Problem are equivalent and we shall treat them as interchangeable.*

2.1 2D rectangular lattice

Remark 1. While a metric ball B_r optimizes the score for triangular folding, that is not true for rectangular folding. Compare the ball B_3 in Figure 6 (which is a ball under the L^1 metric) to the square (or L^∞ ball) of the same size $5^2 = 1 + 4 + 8 + 12$.

Theorem 11. *For each n there exists a word w_n and a fold P_n^{rect} of w_n in the rectangular lattice such that the 0s of w_n form a $(2n+1)$-square and w_n contains no subword 0^{2n+1}.*

Proof. We construct a suitable fold P_n^{rect} of a word starting with $0^{n+1}1$ in a square. See Figure 8 for an illustration of P_5^{rect}. For an illustration of P_2^{rect}, before and after adding whiskers, see Figure 7.

Now we use some classical results:

Theorem 12 ([3, Fact 3.1]). *Among all sets of cardinality n^3 in the rectangular lattice, an n-cube strictly maximizes the number of internal edges.*

Theorem 13. *Among all sets of cardinality n^2 in the rectangular lattice, an n-square strictly maximizes the number of internal edges.*

Proof. The proof is the same as that of Theorem 12.

Theorem 14. *There are infinitely many counterexamples to the* cul-de-sac *conjecture in the rectangular lattice.*

Proof. Let w_n be as in Theorem 11. Assume for contradiction that $J_{\text{rect}}(1w_n1) = J_{\text{rect}}(w_n)$. Thus, there is a fold of $1w_n1$ achieving as many points as w_n. This fold must include a $(2n+1)$-square of 0s by Theorem 13. Since $1w_n1$ starts and ends with 1s, the fold must start and end outside this $(2n+1)$-square. In order to reach the 0 at the center of the $(2n+1)$-square and return outside the square, the word w_n must contain a factor 0^{2n+1}. However, by construction w_n does not contain such a factor.

2.2 3D rectangular lattice

A similar construction succeeds for the 3D rectangular lattice.

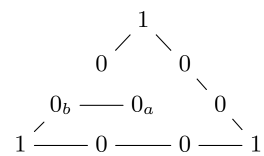
Theorem 15. *For each $n \in \mathbb{N}$ there exists a word w_n and a fold P_n of w_n in the 3D rectangular lattice such that the 0s of w_n form an n-ball and w_n contains no subword 0^{2n+1}. Thereby w_n is a counterexample to the* cul-de-sac *conjecture.*

Proof. Using the 0^3-free length-54 word

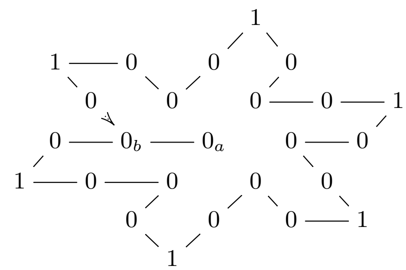
$$(0011)^{12}011100 = 00(1100)^{11}1101(1100)^1$$

a suitable fold is shown in Figure 9.

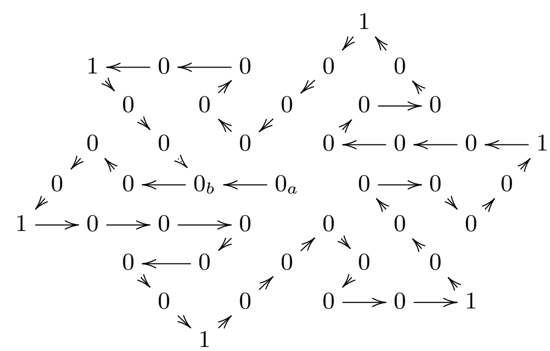
$w_1 = (001)^3 0$ x_1 :



$w_2 = (0001)^6 0$ x_2 :



$w_3 = 0^5 1(0^6 1)^5 00$ x_3 :



$w_4 = 0^5 10^3 1(0^7 10^3 1)^5 0^3$ x_4 :

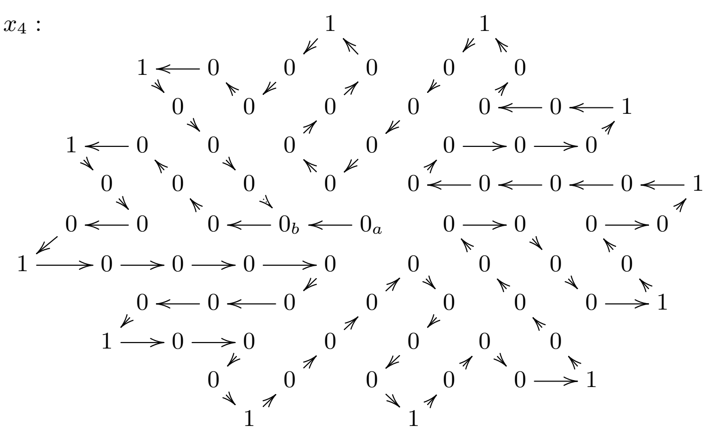


Fig. 4: Folds that demonstrate nonmonotonicity in the triangular model.

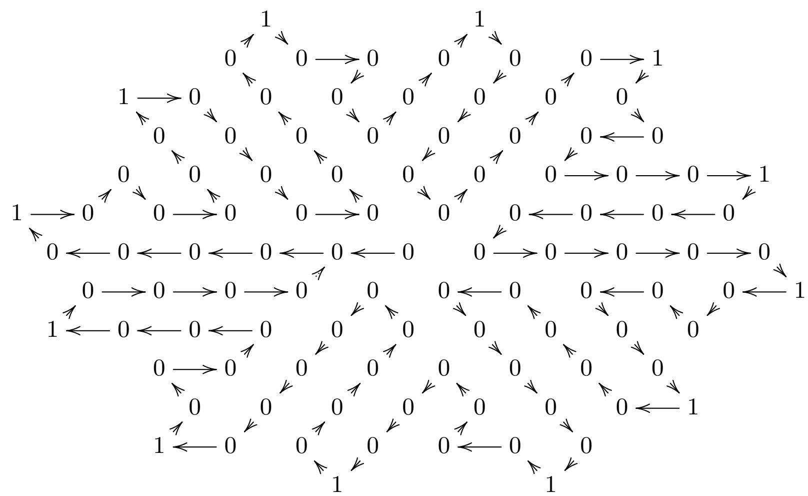
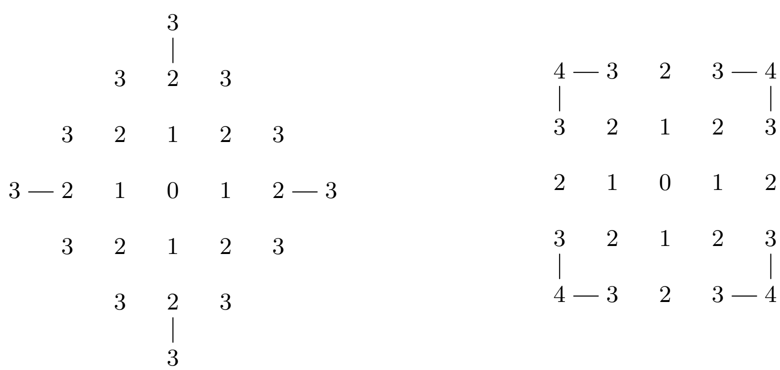


Fig. 5: The fold x_5 in the triangular lattice of a word w_5 having no occurrence of 0^{11}.



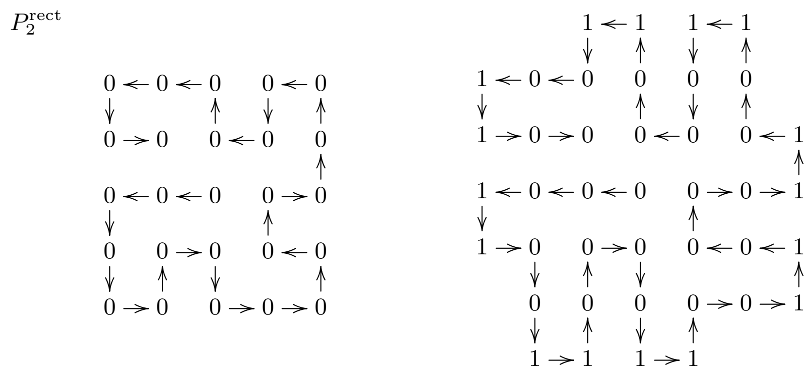
(a) $B_3(L^1)$ has 25 vertices and 36 edges.

(b) $B_2(L^\infty)$ has 25 vertices and 40 edges.

Fig. 6: An optimal configuration (b) in the 2D rectangular lattice that is not a ball (a) in the graph metric. The edges shown are the edges in the symmetric difference between the two graphs.

2.3 Berger-Leighton inspired method

We can also refute the *cul-de-sac* conjecture in an easier way. In Figure 10, the central square is labeled with 0s throughout and the boundary of the square is labeled with 1s. In any optimal fold, the 1s would have to be on the boundary



(a) Before whiskers. (b) With whiskers.

Fig. 7: Folding to obtain nonmonotonicity in the 2D rectangular model.

and completely cover the boundary. Therefore, we have nonmonotonicity in the form $J(w) > J(w1)$ for the word $w = 0^{16}0110((1100)^2 110)^3 1100110$.

3 NP-hardness result

Theorem 16. *The set* $\{(w, k) \mid J_{\text{cube}}(w) = k > J_{\text{cube}}(w1)\}$ *is NP-hard.*

Proof. Recall the following decision problem from Berger and Leighton [3].
MODIFIED BIN PACKING

Instance: A finite set U of items, a size $s(u)$ that is a positive even integer for each $u \in U$, a positive integer bin capacity B, and a positive integer K, where $\sum_{u \in U} s(u) = BK$.

Question: Is there a partition of U into disjoint sets $U_1, \ldots, U_K$ such that the sum of the sizes of the items in each U_i is precisely B?

For each instance it is clear that B is even. Before proceeding with the construction we replace K by $K + 1$, i.e., add one more bin, and add one more item u' with the size $s(u') = B$. When ordering the items we make sure that u' is the last in the sequence. Clearly, in any solution u' would have to be assigned its own bin, without loss of generality be taken to be the last one in the ordering of bins used in Berger and Leighton's construction. It is clear that this modified instance has a solution if and only if the original instance has a solution.

The construction in Berger and Leighton completely covers the surface of the cube with 1s. We can rotate the word so that the long factor of zeros (Hs) is at the ends. That is, where Berger and Leighton's word is of the form $S_{\mathcal{B}} = uH^{(n-2)^3+2}v$, we consider instead $H^a vuH^b$ where $a + b = (n-2)^3 + 2$ and $a, b > 0$.

As the location of the bins is determined by the instance, not by its solution, and since we can add a spurious last item with size exactly the bin size B and

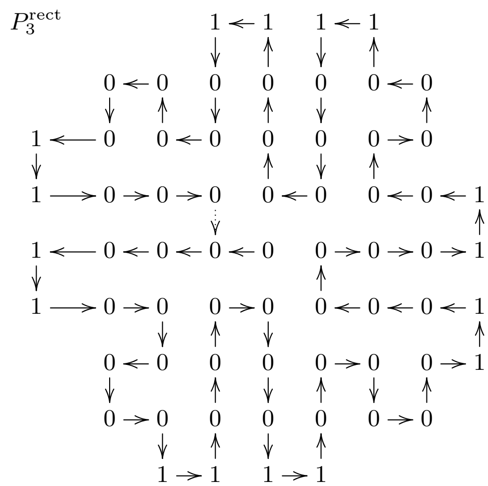


Fig. 8: Folding to obtain nonmonotonicity in the 2D rectangular model.

put it in the last bin, we can control where we are at the end of folding $S_{\mathcal{B}}$ independently of the solution. Then we pad with Ps (1s) and route the fold over to Berger and Leighton's starting node $(n-2, n-1, n-1)$ to complete the loop.

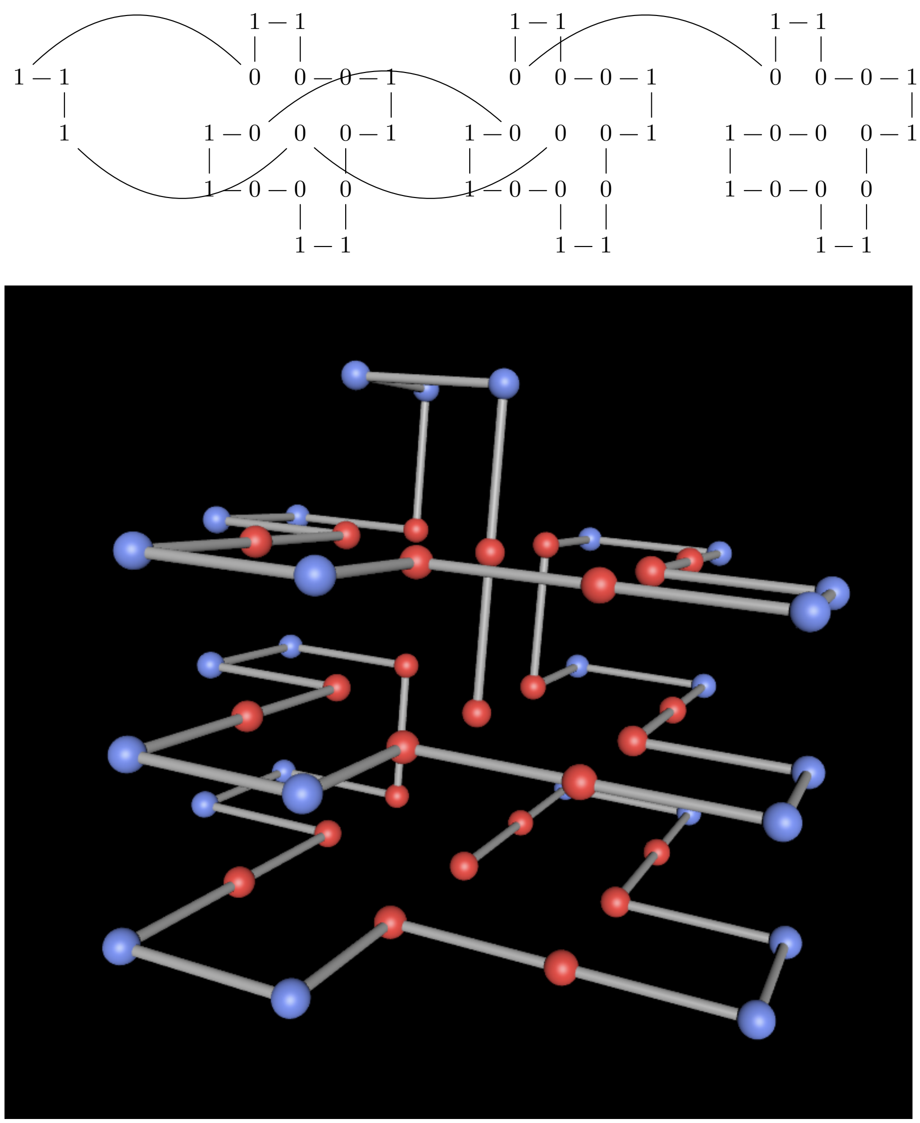


Fig. 9: Folding to obtain nonmonotonicity in the 3D rectangular model for the word $w = (0011)^{12}011100$. For a draggable view of the graphics see https://math.hawaii.edu/~bjoern/?sheets=6&xs=17&ys=6&string=0011001100110011001100110011001100110011001100110011011100&page=labbyfold&moves=aasddsdwwdwaawasewdsddsassawaawdeasddsdwwdwaawasedsff.

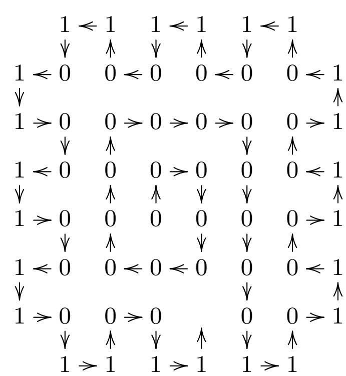


Fig. 10: A Berger and Leighton inspired refutation of the *cul-de-sac* conjecture.

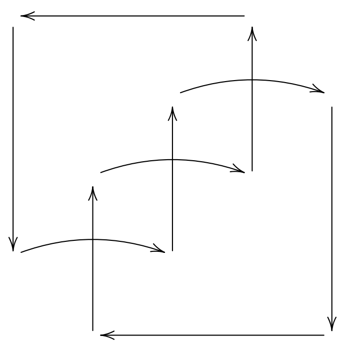
4 Knots and links

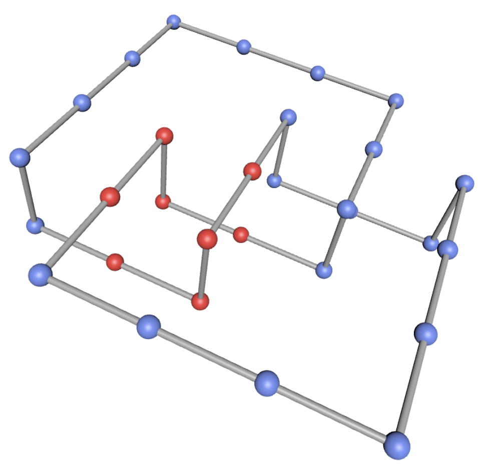
We do not know whether knots are necessary for the optimal value function J_{cube}. Wüst, Reith and Virnau [13] considered knotting in the HP model, but only from a statistical point of view, hence they do not address whether knotting is necessary for optimal folding.

Of course, an optimal closed fold can include any knot, simply by folding a word of the form 1^* in the shape of that knot. Here a *closed* fold is a walk in the lattice that is self-avoiding except that the first and last vertices coincide.

Let us therefore say, given a knot K, that a *proper K-fold* of a word w is a closed fold of w which looks like K (i.e., is ambiently isotopic to K) and such that in a 2-dimensional knot diagram of the fold, the over-and-under vertices are separated by points scored with respect to w.

Here is a diagram of the trefoil knot with over-and-under vertices indicated by curving.

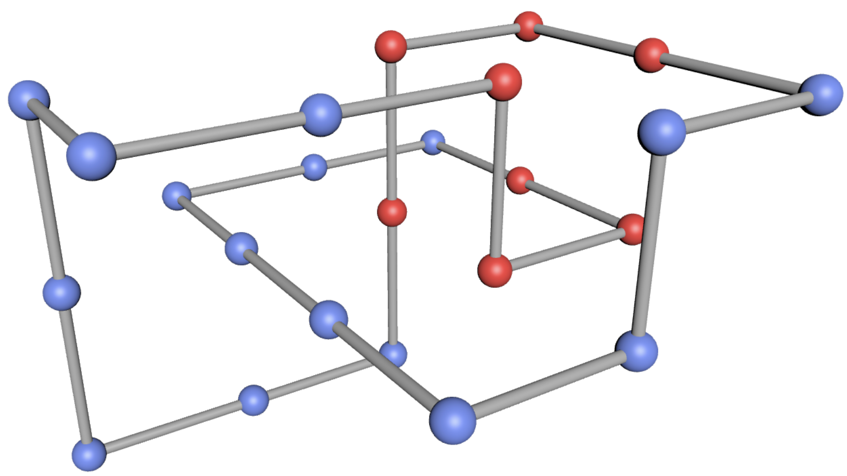




It is immediate that only if K is *the unknot* can there be a proper K-fold of a word of the form 1^*.

Theorem 17. *There exists a cyclic word w, a nontrivial knot K, and a proper K-fold x of w such that x is an optimal fold of w.*

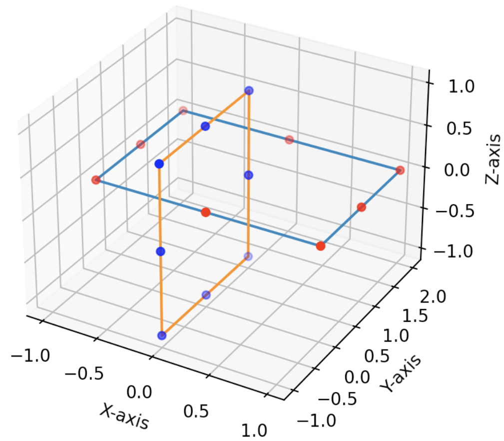
Proof. Let $w = 1^2(0^4 1^7)^2$, a word of length 24, let K be the trefoil knot, and let x be the fold shown below. As seen earlier, the central red cube (where red is identified with 0 and blue with 1) guarantees an optimal fold.



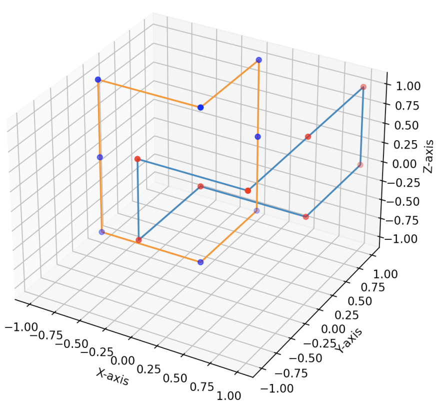
```
https://math.hawaii.edu/~bjoern/?sheets=9&xs=17&ys=17
&string=110000111111100001111111
&page=labbyfold&moves=eddfwwffssaeeweddffaaaes
```

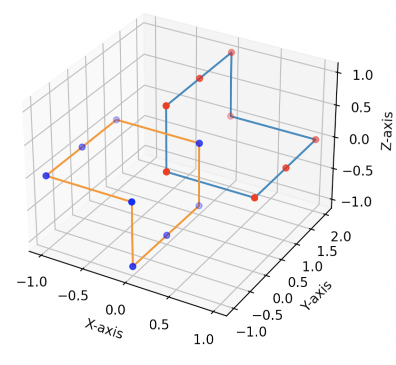
It is clear that we can achieve a central packed cube of 0s with a minimal length knot. The minimal length of a trefoil knot in the 3D integer lattice is 24 [9]. This was already achieved above for the word $1^7 0^4 1^9 0^4$ and 6 points. However, this is not uniquely optimal as we can fairly easily take the given configuration of red monomers and add 7 and 9 blue monomers (using 8 and 10 edges) in a non-knotted manner. Thus, we know that a knot can appear in a minimal-length, optimal-score, but not-uniquely-optimal-score, manner in the 3D HP model.

We have seen that knots do occur as *optimal* folds, even if we have not established that they occur as *uniquely optimal* folds. For links the situation is no easier. Consider the "polymer complex" consisting of two copies of $(01)^4$. Thus, we have two loops of length 8 labeled $(01)^4$. If we link them so that 0s are in the middle(s), we achieve 7 points:

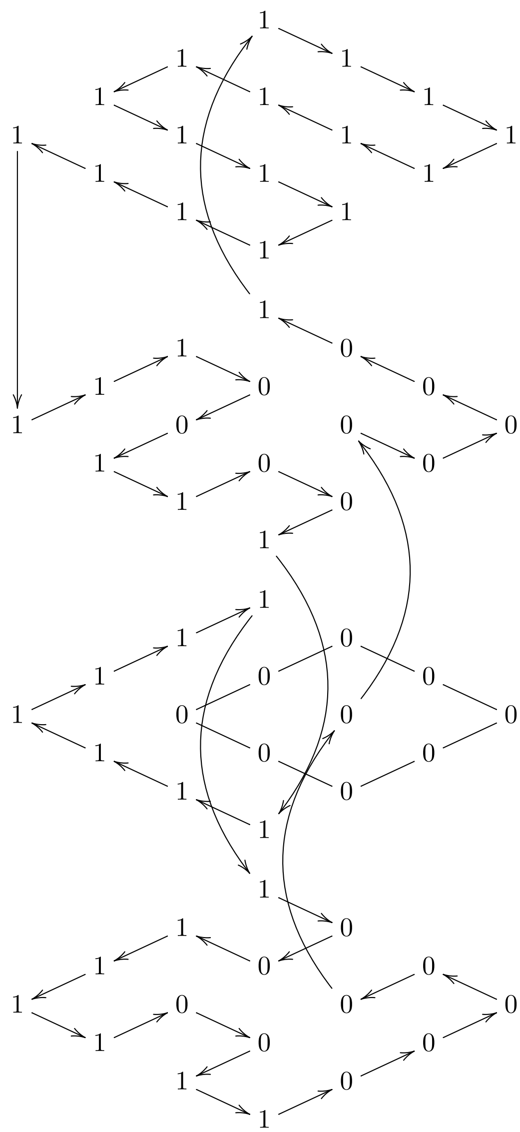


We can show by brute force search that links can be optimal, even if we cannot show they are uniquely optimal. Behold examples for two copies of $(01)^4$:

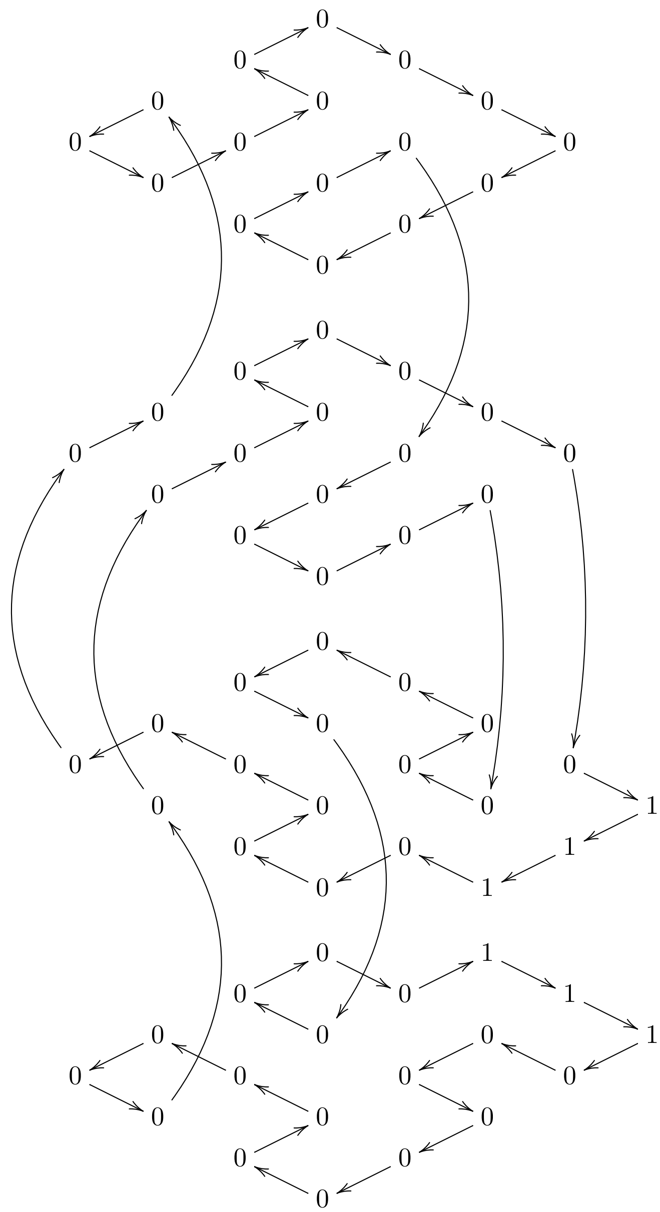




An optimal link. Below we have a 4-cube formed from two linked loops of length 8 and $64 - 8 = 56$, respectively. We can make every monomer a 0 (hydrophobic) but it is more interesting to only make a certain 3-cube that way. Then the words read are 0^8 and (starting from the linking point) $0^6 1^{20} 0^2 1^2 0^2 1^9 0^2 1^4 0^2 1^2 0^5$. It is not hard to see that these two words can also form a 4-cube in an unlinked way, however.



An optimal trefoil knot. Since the 0s form a $4 \times 4 \times 4$ grid, the following is an optimal fold of the length-70 cyclic word $1^3 0^{22} 1^3 0^{42}$ in the shape of a trefoil knot.



Let us say that the *hydrophobic core* of a fold is the map sending locations in the word to locations of the hydrophobic monomers, that is undefined on polar monomers. Let us moreover say that an *essential fold* of the knot K is a fold such that all folds with the same hydrophobic core are homotopic to K. Then the above example shows that an essential fold of the trefoil knot 3_1 can be an optimal fold of a word. We conjecture that this is true for all knots.

Acknowledgments

This work was partially supported by grants from the Simons Foundation (#704836 to Bjørn Kjos-Hanssen) and Majesco Inc. (University of Hawai‘i Foundation Account #129-4770-4).

References

1. Richa Agarwala, Serafim Batzoglou, Vlado Dančík, Scott E. Decatur, Martin Farach, Sridhar Hannenhalli, S. Muthukrishnan, and Steven Skiena. Local rules for protein folding on a triangular lattice and generalized hydrophobicity in the HP model. In *Proceedings of the First Annual International Conference on Computational Molecular Biology*, RECOMB '97, page 1–2, New York, NY, USA, 1997. Association for Computing Machinery.
2. Oswin Aichholzer, David Bremner, Erik D. Demaine, Henk Meijer, Vera Sacristán, and Michael Soss. Long proteins with unique optimal foldings in the H-P model. *Computational Geometry*, 25(1):139–159, 2003. European Workshop on Computational Geometry - CG01.
3. Bonnie Berger and Tom Leighton. Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. In *Proceedings of the Second Annual International Conference on Computational Molecular Biology*, RECOMB '98, page 30–39, New York, NY, USA, 1998. Association for Computing Machinery.
4. Pierluigi Crescenzi, Deborah Goldman, Christos Papadimitriou, Antonio Piccolboni, and Mihalis Yannakakis. On the complexity of protein folding. *Journal of Computational Biology : a journal of computational molecular cell biology*, 5:423–65, 02 1998.
5. Elisa Davoli, Paolo Piovano, and Ulisse Stefanelli. Sharp $N^{3/4}$ law for the minimizers of the edge-isoperimetric problem on the triangular lattice. *J. Nonlinear Sci.*, 27(2):627–660, 2017.
6. Ken A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985. PMID: 3986190.
7. Matthew T. Gilzinger. Exploring the HP model for protein folding. Technical report, Worcester Polytechnic Institute, 100 Institute Road, Worcester MA 01609-2280 USA, April 2012.
8. L. H. Harper. *Global methods for combinatorial isoperimetric problems*, volume 90 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2004.
9. Kyungpyo Hong, Hyoungjun Kim, Seungsang Oh, and Sungjong No. Minimum lattice length and ropelength of knots. *J. Knot Theory Ramifications*, 23(7):1460009, 10, 2014.
10. John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Kathryn Tunyasuvunakool, Olaf Ronneberger, Russ Bates, Augustin Zídek, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Anna Potapenko, Andrew J Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Martin Steinegger, Michalina Pacholska, David Silver, Oriol Vinyals, Andrew W Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. High accuracy protein structure prediction using deep learning. In *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book)*, CASP, 2020.

11. Bjørn Kjos-Hanssen. *Automatic Complexity: A Computable Measure of Irregularity*. De Gruyter, Berlin, Boston, 2024.
12. Jack Stecher. Personal communication, March 2023.
13. Thomas Wüst, Daniel Reith, and Peter Virnau. Sequence determines degree of knottedness in a coarse-grained protein model. *Phys. Rev. Lett.*, 114:028102, Jan 2015.