Topology-Free Folding Pathways of Polyhedral Nets

Paul M. Dodd[1], Pablo F. Damasceno[2], & Sharon C. Glotzer[1]

[1]*Chemical Engineering Department, University of Michigan, Ann Arbor, Michigan 48109, United States*

[2]*Department of Cellular and Molecular Pharmacology, University of California, San Francisco 94158, United States*

Understanding the relationships between the ~~properties of 1 and~~ geometrical properties of 1-dimensional objects such as proteins or 2-dimensional objects such as sheets and their folding propensity remain a major challenge for *de novo* design of 3-dimensional objects. Using ~~Molecular Dynamics~~ molecular dynamics simulations, we investigate the process by which the three simplest Platonic solids can be folded ~~from their twenty four 2-dimensional~~ via thermodynamic self-assembly from their 24 two-dimensional regular nets. We show that ~~within one~~ for a given target folded shape, ~~higher~~ the number of leaves, ~~higher smoothness and higher~~ the number of traps accessible to the unfolded state, and the number of pathways connecting unfolded to folded state ~~positively~~ correlate with folding propensity. We demonstrate that the nets follow a ~~topology-independent mechanism governs the nets folding process, following a~~ universal balance be-

tween freezing of internal degrees of freedom and potential energy gain as expected for processes in or near-equilibrium. ~~Finally, following~~ Exploiting this mechanism, we devise a numerical method to enumerate all equilibrium folding pathways. ~~allowing~~ Allowing for an efficient evaluation of the folding of the more complex Platonic solids, against which the correlations above are tested~~against~~. Our results ~~present an~~ provide a heuristic for the design of 2-dimensional objects able to successfully fold into target 3-dimensional geometries of interest.

~~**intro from talk combine with one that has already been written** Colloidal particles receive much attention today because of the wide variety of structures that they can form, and the immense amount of control with which they can be synthesized. What makes the particles interesting is the anisotropic features that can be designed to correspond to specific targeted assemblies. And we are truly coming into the age of designer particles that could be ordered on demand and made by the bucket. In these hard polyhedra systems the shape of the particles, such as the sphericity, determines which broad class of assembly each particle assembles into. One major problem is that some (or most) of these particles are not able to be crystalized because the angles of the facets would be geometrically forbidden for atoms to crystalize in. But already we can ask the next question of "how would one make some of these particles?". As you may have guessed from the title of~~

~~the talk we are proposing to use folding. What you may not have guessed is that this question was asked nearly 500 years ago by Albrect Durer. Arguably, the first person to think about unfolding polyhedra was Albrect Durer, he was the first person to publish this technique. He was Renaissance man from 1500's. Durer was an artist as well as a mathematician, and wanted to build these symmetric shapes for his artwork. He realized that to make them out of clay or other media was very difficult but he discovered that if he started with a unfolded form, what he calls a net, then it was much easier to fold the shape later. So why not use the exact same approach for particles that are 1 million times smaller.~~

Self folding origami has been an emerging field with the goal to design materials that can self-assemble with applications ranging from drug delivery[?] to robotics[?]. Many current studies have shown how to leverage physical response to light[?], heat[?], pH[?], and stress[?] to achieve controllable folding using polymer sheets[?], silicon wafers[?], cellular scaffolding[?], and DNA[?].

~~If you want to know how to fold say a cube the first thing you need to do is unfold it.~~ In this work we study a specific type of unfolding, called edge-unfolding. This is where you are only allowed to cut along the edges of the shape in such a way that you don't cut it into more than one piece and the result can lay flat in the plane. ~~There are only 5 regular polyhedra. There are 2 nets for the tetrahedron, 11 nets for~~

~~the cube and octahedron, and 43,380 for the dodecahedron and icosahedron.~~

~~David Gracias studied self folding on the millimeter length scale. Recognizing that if you want these facets or plates to come together to form a cube if each one is free in solution there are many degrees of freedom, but if you can tether some of the faces together then you can reduce the total number of DOF's in your system and assembling the shape becomes much more feasible. Due to the length scale of these nets the folding is deterministic, angle between the hinges is determined by how much the solder is placed on each hinge, and irreversible. After the hinge has folded there is no going back.~~ The main finding that Gracias and his colleagues found is that compact nets fold better. ~~Now we ask what will happen if the nets are allowed to fold due to purely stochastic forces?~~

The combinatorial explosion in the number of nets means that we cannot even simulate all of the nets for the icosahedron and dodecahedron so we need to first establish which nets fold best when all of the folding is occurring randomly.

Understanding the process by which lower-dimensional geometries fold into 3-dimensional structures is important in many fields. In the 16th century, the Dutch artist Albert Durer investigated which ~~2-dimensional~~ 2D cuts of non-overlapping, edge-jointed polygons could be folded into an ~~platonic and archimedean polyhedron~~Platonic

and Archimedean polyhedra[?,?]. Mathematically, those cuts were later defined as "nets" and many questions such as whether any convex polyhedron can be decomposed into a net remain open[?,?,?]. In natural systems, the canonical example of folding ~~happens~~ occurs in proteins, RNA and DNA. There, 1-D strings of molecules folded from a denatured (unfolded) state to a natured (folded) state. **include foldamers – Dill**

In larger scale artificial systems, this relation between geometry and folded configuration can be investigated with higher control. In the macroscopic folding of kirigami sheets, for example, ~~relations~~ relationships between cuts/creases and macroscale properties such as elongation have been ~~stablished~~established [?,?]. In fact, the reverse problem of finding cuts leading to a particular geometry, even allowing for the pluripotent design of sheets able to fold into multiple states~~have~~ , has also been demonstrated[?]. Other experimental systems have explored the use of ~~use~~ light[?], capillary forces[?,?], pH[?], cellular traction[?], and thermal expansion[?] to achieve controllable self-folding behavior on the macroscale.

~~In smaller scale~~On smaller scales, folding of ~~microscopic~~ polyhedral capsules have ~~allowed for the discovery of~~ revealed relationships between geometry and folding yield[?] but the relatively large scale of the building blocks to the length scales of the solution in which the capsules were contained, resulting in a deterministic fold-

ing process. However, if the structures are made much smaller – below a micron – then the folding process will ~~become~~ be stochastic. Our aim is to understand how these ~~origami-like structures~~ 3D objects can be formed through the use of stochastic forces and origami-like folding processes.

In this study, we investigate by means of Langevin ~~Dynamics~~ molecular dynamics simulation the stochastic folding of regular nets of the ~~5~~ five Platonic solids: both nets for the ~~tetrahedra~~tetrahedron, all 22 nets ~~for cube and octahedra~~(11 each) for the cube and octahedron, and 80 out of the possible 86,760 nets ~~for the dodecahedra and icosahedra~~(43,380 each) for the dodecahedron and icosahedron. By doing so, we investigate the ~~process~~ thermodynamic and kinetic processes by which different net geometries fold into a pre-designed, common ground state.

The basic design ~~cycle that~~ approach we are using to find the best nets – that is, those with the highest folding propensity – is as follows: first we start with the target shape, we translate this shape into a graph representation and using some mathematics we can enumerate all of the topologically distinct nets. We then translate these nets into a molecular dynamics model. Finally, we run simulations to measure and observe the folding and compare the final shape in our simulation to the target structure to quantify the yield.

~~**move methods later?**~~ ~~Using Lengevin Dynamics~~ Using Langevin MD in HOOMD-Blue, we approximate each face as a union of spheres that act as a rigid body. The blue ~~sphere exhibit~~ spheres interact via a purely repulsive WCA potential and the gray spheres ~~exhibit~~ interact via the Lennard-Jones potential~~. The~~, which includes both repulsion and attraction. The blue and gray spheres interact repulsively. The faces are tethered together using harmonic springs. Details of the simulation model, methods and procedures are in the SI. All of the nets for the tetrahedron, cube and octahedron are shown in figure XX in the SI.

~~**move methods later?**~~

1 Results

To quantify the folding yield we ~~run~~ ran 125 simulations starting from a high temperature and quench the temperature to ~~a really low temp~~near zero. We then define the yield as the fraction of simulations that folded. ~~And we~~ We define the folding propensity as the average number of native contacts averaged over all of the quenching simulations we ~~run.~~ ran. If all nets folded perfectly in all runs, the folding propensity would equal one.

We find that there is one tetrahedron net that folds 100% of the time and one

that folds around 50% of the time in the fastest quench we try. We find that in general there are only a few nets that will reliably fold into the cube and octahedron. We find that XX of the nets for the octahedron can fold into another type of concave polyhedron, the poly-tetrahedron. This same sort of structural competition has been seen in clusters of attractive colloids, where symmetry breaking leads to a higher rotational entropy [?]. While the potential energy of the poly-tetrahedral conformation and the octahedron are the same, on the free energy landscape the poly-tetrahedral is the minimum ~~for this reason. One can design away these extra potential traps by changing the topology of the net and in general the net with the most funneled landscape will give the best yield.~~ because it has higher rotational entropy. For these nets the octahedron is actually metastable to the poly-tetrahedron. **One can design away these extra potential traps by changing the topology of the net and in general the net with the most funneled landscape will give the best yield. – see hard copy**

~~**move methods later?**~~ To understand more deeply why different nets fold with different probabilities~~. We compute~~ , we computed the folding pathways using Markov ~~State models~~ state models (MSM). We define the states using two order parameters, the bonds and the dihedral angles of each hinge edge. We ~~can use~~ used Transition Path Theory [?,?] to determine the reactive flux, $f_{ij} = q_j^+ \pi_i P_{ij} q_i^-$,

along the dominant pathways. The dominant paths ~~are~~ were computed using the algorithm from ~~?~~ ?,?, see the SI for ~~more~~ details. To build the MSM we ~~runs~~ ran NVT simulations for 10^9 steps~~then branch~~ , and then branched the simulation until 14 total simulations ~~are obtained~~ were obtained (or an equivalent of (??) timesteps~~.~~ ~~We sample~~). We sampled our order parameters every 10 timesteps~~,~~ ; the lag time was found in the usual way of finding the time at which the eigenvalues of the P matrix become constant ~~as seen in Figure SIXX~~(Figure SIXX).

After calculating the reactive flux we can plot the folding network for each net~~. We see that that~~ , we see that there are two ~~main pathways that~~ pathways along which the linear net tetrahedron can fold. One pathway uses the only intermediates that have native contacts (contacts that are also in the folded structure). The other main pathway uses ~~native contacts. We observe that the relative flux through pathways has a temperature dependence. At low temperatures the nets tend to fold through pathways that have~~ non-native contacts ~~, as the temperature increases there is a transition to use pathways that only use native contacts. This sort of behavior is seen in other systems where on one side there is a large entropic basin, and the other there is a large potential energy basin??. Another connection to proteins is that several small proteins will fold via native only contacts at the melting temperature ??,??,??. As the proteins have increasing hydrophobicity or~~

~~lower temperature the pathways shift to a hydrophobic collapse of the chain and then further rearrangement to eventually fold into the native state?.~~

(bonds that must break before the net can fold completely). Non-native contacts play either an "active" role or a "passive" role in the folding process. The "active" role is when a native contact is brought together by the non-native contact. In general, this happens when a nonlocal bond needs to be formed and the non-native contacts stabilize the structure. The more common role is for the non-native contact is passive. This is when a non-native contact is formed and then the net folds using native contacts, as if the non-native contact does not exist, at the end the non-native contact breaks and the net can fold into the final target structure.

We observe that the relative flux through pathways has a temperature dependence. At low temperatures the nets tend to fold through pathways that have non-native contacts. As the temperature increases there is a transition to use pathways that only use native contacts.

When calculating the folding yield the system is coming from high temperature during the quenching procedure, so the pathways are likely to be dominated ~~using these pathways~~ by pathways that use native contacts only. A representative high temperature pathway for the cube and octahedron nets is shown in figure 4. We

see that the pathways in the high temperature limit are *cooperative*, meaning that they fold using local bonds first, which in turn bring non-local contacts that were farther away close together ~~which folds~~ that fold next. We observe that as each net folds it will freeze out the fewest number of degrees of freedom as possible; therefore, maximizing the conformational entropy along the folding pathways. Because of this ~~affect~~ effect of entropy we find different dominant intermediates than previously observed by Gracias et al. [?,?], in a deterministic study of the same folded structures. In our model entropy is being maximized, where their model is deterministic so the pathways they observe will follow the principal of least action.

To further illustrate ~~this~~ the trade off between the degrees of freedom and native contacts at high temperature, we project all of the pathways (both native and non-native) onto these two order parameters: the normalized number of degrees of freedom N and the number of native contacts Q, the results are shown in figure 5. The size of each data point corresponds to the likelihood that any net will fold along a path that has any intermediate with those order ~~parameters~~parameter values. The curves in figure 5 are the weighted ~~average~~ averages of N holding Q constant. The lines are drawn to guide the eye. As shown, each net folds along these two order parameters in the same way. The upper right corner of the plot is a geometrically forbidden region arising from the fact that the net can not gain a bond without losing

at least one degree of freedom. ~~The sequential (cooperative) pathways agree with the proposed theory that small proteins fold using foldons, which are subsections of the protein that form a secondary structure. The protein can have many foldons that fold in a certain order. One foldon will not form before its predecessor has already folded?**make sure this is solid**. This model of protein folding seems to agree with the idea that the unfolded part of the protein will want to maximize entropy as the other parts start to fold. The other observation is that the protein undergoes a rigid to non-rigid transition?, which clearly occurs in these nets as well. It was also observed that in unfolding simulations, the proteins unfolded about a rigid core .~~

There are, in general, many different ways that a net can fold along the two order parameters N and Q. The number of these pathways accessible to the net will cause the net to have a higher folding yield, coming from the fact that ~~$k \propto N$. The folding~~ $k \propto N_{paths}$. Folding will always start at a vertex that corresponds to a leaf (degree one vertex) on the cutting tree that we computed for the net enumeration. So yield should also correlate with the number of these vertices on the net as well. We show that the yield correlates to these parameters for the cube and octahedron in the first two columns of Figure 6. We therefore use these parameters to make predictions about which of the dodecahedra and icosahedra nets will fold best and worst. These predictions are shown in figure SI-XX. We find that these two param-

eters are good predictions and can be used as for design rules for future nets.

~~**Think about the ordering**~~

2 Discussion

This sort of behavior is seen in systems of patchy particles where on one side there is a large entropic basin, and the other there is a large potential energy basin[??]. Another connection to proteins is that several small proteins will fold via native only contacts at the melting temperature, where the protein is denatured 50% of the time[??,??,??]. As the proteins have increasing hydrophobicity or lower temperature the pathways shift to a hydrophobic collapse of the chain and then further rearrangement to eventually fold into the native state[?]. This is analogous to using pathways with non-native contacts.

The sequential (cooperative) pathways agree with the proposed theory that small proteins fold using foldons, which are subsections of the protein that form a secondary structure. The protein can have many foldons that fold in a certain order. One foldon will not form before its predecessor has already folded[?] ~~**Add a conclusion**~~**make sure this is solid**. This model of protein folding seems to agree with the idea that the unfolded part of the protein will want to maximize entropy

as the other parts start to fold. The other observation is that the protein undergoes a rigid to non-rigid transition[?], which clearly occurs in these nets as well. It was also observed that in unfolding simulations, the proteins unfolded about a rigid core **double check this**[?].

Acknowledgements Put acknowledgements here.

Competing Interests The authors declare no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to S.C.G.(email: sglotzer@umich.edu).

1. ~~Motivation. some nets are pluripotent. can fold into different ground states.~~
2. ~~flush out the methods sections in full detail.~~
3.

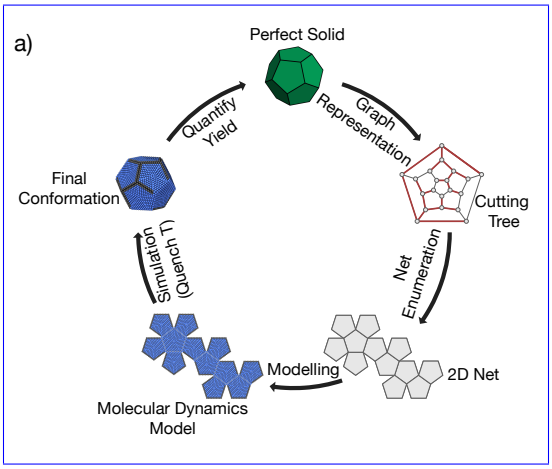

Figure 1: a) Example workflow used in this study. Starting with the perfect solid, the target shape, the graph representation is used to enumerate all the different ways to cut the solid. The vertices in the graph correspond to the vertices of the polyhedron and the edges in the graph correspond to the edges of the polyhedron. The edges marked in red represent the edges that are cut on the shape. The all red edges form a spanning tree for the solids studied in this work. Using the information in the cutting graph we obtain the 2-dimensional net of the polyhedron. The proposed model to simulate the folding process. Each face is modeled as a union of spheres held rigidly together. Faces are connected via harmonic bonds. The simulations are initialized at high temperature and then the temperature is quenched to low temperature. Once the final temperature is reached the final configuration is compared to the target shape and the folding probability (yield) is calculated. b) the pairwise interactions used in the molecular dynamics simulations. The blue-blue and blue-grey interactions in the model exhibit the Weeks-Chandler-Anderson (WCA) potential (blue line) and grey-grey interaction in the model exhibit the Lennard-Jones(LJ) potential (gray line). (c) Folding yield defined by the fraction of simulations that reached the folded state for the tetrahedron, cube, octahedron, and 11 of the 43,380 dodecahedron nets. The folding yield is dependent on the geometry of the net. Generally the more compact nets fold better and nets that have a high number of leaves on their cutting tree and a smaller diameter usually folds best.

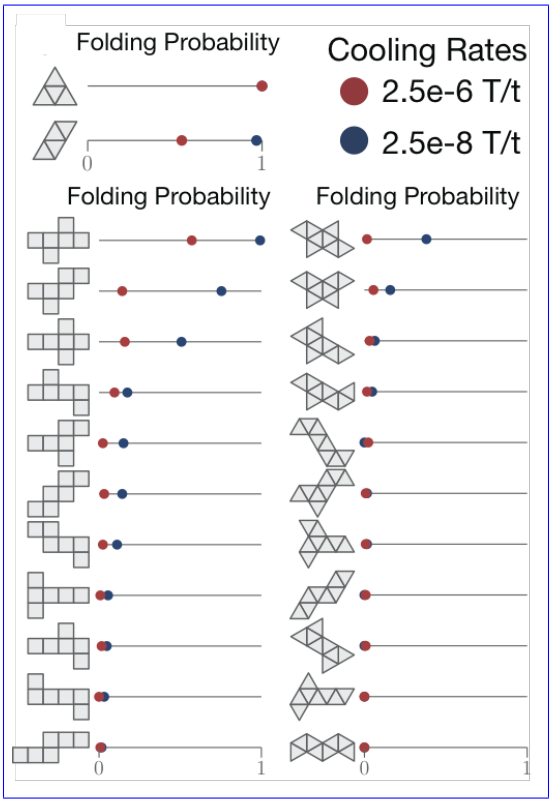

Figure 2: a) Example workflow used in this study. Starting with the perfect solid, the target shape, the graph representation is used to enumerate all the different ways to cut the solid. The vertices in the graph correspond to the vertices of the polyhedron and the edges in the graph correspond to the edges of the polyhedron. The edges marked in red represent the edges that are cut on the shape. The all red edges form a spanning tree for the solids studied in this work. Using the information in the cutting graph we obtain the 2-dimensional net of the polyhedron. The proposed model to simulate the folding process. Each face is modeled as a union of spheres held rigidly together. Faces are connected via harmonic bonds. The simulations are initialized at high temperature and then the temperature is quenched to low temperature. Once the final temperature is reached the final configuration is compared to the target shape and the folding probability (yield) is calculated. b) the pairwise interactions used in the molecular dynamics simulations. The blue-blue and blue-grey interactions in the model exhibit the Weeks-Chandler-Anderson (WCA) potential (blue line) and grey-grey interaction in the model exhibit the Lennard-Jones(LJ) potential (gray line). (c) Folding yield defined by the fraction of simulations that reached the folded state for the tetrahedron, cube, octahedron, and 11 of the 43,380 dodecahedron nets. The folding yield is dependent on the geometry of the net. Generally the more compact nets fold better and nets that have a high number of leaves on their cutting tree and a smaller diameter usually folds best.

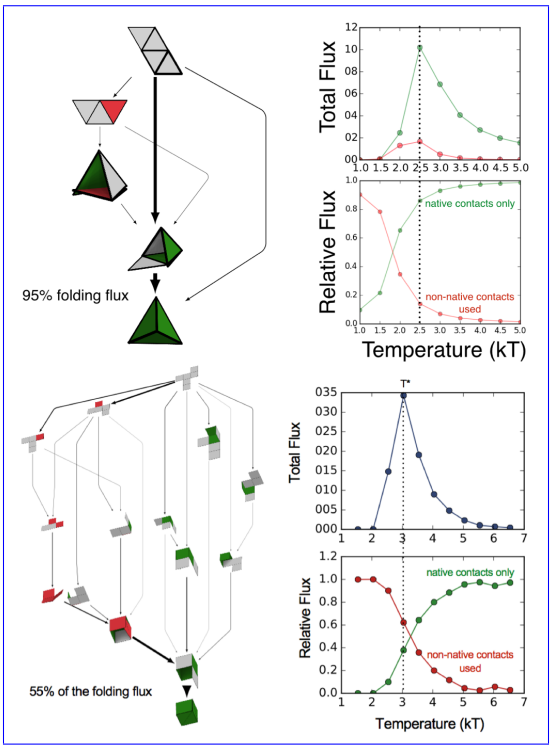

Figure 3: (a) Folding pathways for one of the cubic nets. The network can be separated into two regions: native contacts only(left) and non-native containing (right). The pathways that contain only native contacts follow a "sequential" pathway where one face folds at a time. The non-native pathways are observed to behave in two different ways i) they can bring 2 edges that need to bind close together and later the non-native bonds break leaving the native contact ii) the polyhedron folds as if the non-native contact was not there and later this non-native contact breaks and the folding can complete to the final target structure. (b) The total probability flux of all the pathways. (c) The relative amount of flux going through pathways that use the native-exclusive (green curve) pathways and the non-native inclusive pathways (red pathways).

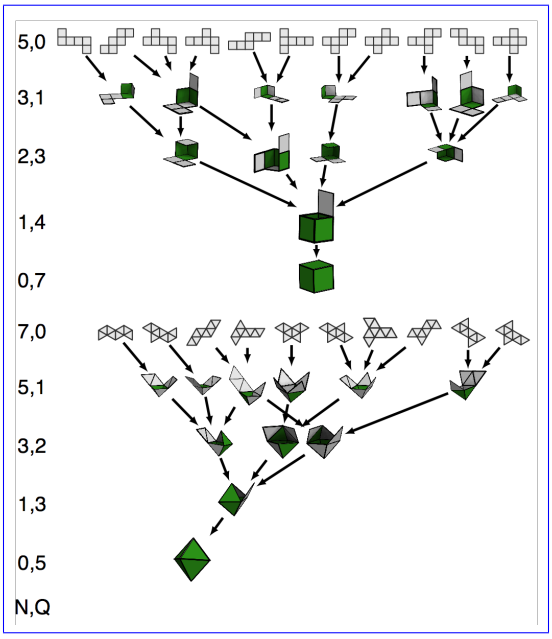

Figure 4: Combined dominant pathways (a) the tetrahedron (b) the cube (c) the octahedron taken at a temperature $T > T^*$. In both a-c we observe that the dominant pathways are sequential, generally speaking one face folding at a time. All of the observed pathways include only native contacts. If we project ALL the pathways of ALL the nets onto the two order parameters native contacts and number of internal degrees of freedom, we find that each net will trade their degrees of freedom in approximately the same way. This is demonstrated for the tetrahedron in (d) the cube in (e) and octahedron in (f).

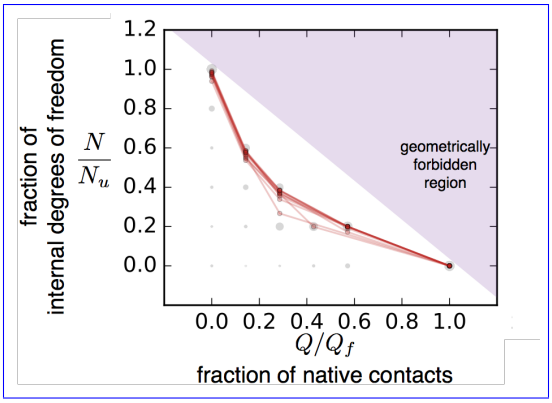

Figure 5: Combined dominant pathways (a) the tetrahedron (b) the cube (c) the octahedron taken at a temperature $T > T^*$. In both a-c we observe that the dominant pathways are sequential, generally speaking one face folding at a time. All of the observed pathways include only native contacts. If we project ALL the pathways of ALL the nets onto the two order parameters native contacts and number of internal degrees of freedom, we find that each net will trade their degrees of freedom in approximately the same way. This is demonstrated for the tetrahedron in (d) the cube in (e) and octahedron in (f).

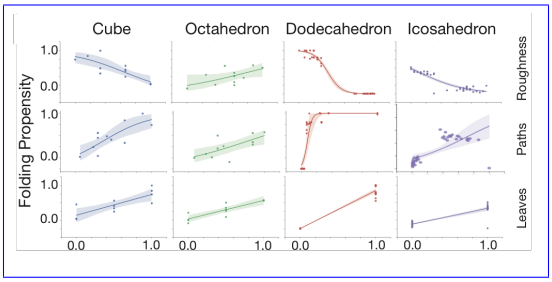

Figure 6: (a) The average fraction of native contacts observed over 125 simulations at the slowest cooling rate for each net. Cubic nets are shown in red and octahedron nets are shown in blue. (will add the other data) There is a strong correlation between the average fraction of native contacts and the folding probability. The average fraction of native contacts helps distinguish poor folders. (b) The average fraction of native contacts vs the number of high temperature pathways shown in Fig. 5a-c. Folding strongly correlates with the number of pathways geometrically accessible to the net. (c) The average fraction for the tetrahedron, cube, octahedron, and dodecahedron. Folding positively correlates with the number of leaves. (d) The average fraction for the tetrahedron, cube, octahedron, and dodecahedron. Folding negatively correlates with the diameter of the net. (e) Common folding motifs that correspond to the leaves on the cutting tree. (show leaves and edges that will fold together, also show the corresponding folded structure) The number of internal degrees of freedom in the folded structure. As seen the octahedron and icosahedron have folded structure that contain internal degrees of freedom. These internal degrees of freedom can lead to traps and in general disrupt the folding. It is observed that the shapes that have rigid folded structures have nets that fold with higher probability than those that have internal degrees of freedoms.