Molecular tinkertoys and how to assemble them

Warren D. Smith*
wds@research.nj.nec.com

May 3, 1997

Abstract — **We propose a systematic method of synthesizing robust macromolecules of any desired shape with precise and total structural control. It involves assembly of 3D objects from a few fundamental and carefully designed "molecular building blocks" in a "treelike leaf-to-root" fashion, followed by a later "global rigidizing" reaction.**

We propose three different sets of abstract primitive operations, (which the building block molecules must support) and show that each suffices for essentially universal synthetic power. We present a mathematical theory of "assembly" including a polynomial time algorithm to find an "optimal" (e.g. with maximum possible synthetic yield) "tree sliced iso-oriented" assembly of any "lattice animal." Finally, we present paper designs of actual molecules and chemistry to show that all of our required primitive operations should indeed be achievable compatibly. Analysis suggests that the major limit on all this will be imperfect chemical specificity.

This idea is still in its early stages and will need further investigation and development, especially by synthetic organic chemists, to create building blocks with the right properties.

This method may provide a way to achieve something similar to K.Eric Drexler's notions of "nanosystems" (popularized in several books authored or coauthored by Drexler). Although I think much of the "nanosystems" area is closer to science fiction than to science, the ideas in the present paper, especially if it is possible to "raise" them to "higher levels" of the "size heirarchy," may provide a way to realize some science fiction visions with some resemblance to Drexler's. In particular, perhaps they may eventually lead to ultrafast and/or ultrasmall computers. In the nearer term, these techniques should lead to micromold or microstencil techniques for creating extremely small structures, and the creation of small quantities of: photonic bandgap materials, highly controllable "molecular sieve" materials, and zero density solids.

Keywords — Nanosystems, dendrimers, molecular sieve, low density solids, micromolds, microstencils, photonic bandgap, molecular tinkertoys, automated synthesis.

Contents

1 The problem **2**

2 The plan in brief **3**

3 Plan of the rest of the paper **4**

4 Comparison with previous work **4**
4.1 Proteins . . . 4
4.2 Dendrimers . . . 5
4.3 "Tectonics"? . . . 7
4.4 The big and small pictures . . . 8

5 Building blocks ("tinkertoys") and spaceframe **8**
5.1 Only three arms are needed, and spaceframes that prove it. . . . 8
5.2 The need for porosity; the unimportance of size limits . . . 9

6 Assembleability **10**
6.1 "Useful" shapes . . . 12
6.2 Polynomial time algorithm to find the best "iso-oriented tree sliced assembly" of a lattice animal . . 12

*NECI, 4 Independence way, Princeton NJ 08544.

7 Sets of Primitive operations **13**
7.1 A set of primitives with universal assembling power based on stacked protection – only one kind of building block needed 13
7.2 Alternative set of universal assembling primitives based on building blocks with DNA arms 14
7.3 Other possible primitive operations 16
7.4 "Directed tree" viewpoint removes the need for stacked protection 16

8 Molecular rigidity and stability **17**
8.1 Rigidity and its limits 17
8.1.1 Atropoisomers 18
8.1.2 Conclusion 19
8.2 Stability 21

9 Some useful molecules and reactions **22**
9.1 A few molecules with interesting geometrical properties 22
9.2 Brief review of some cyclization reactions 23

10 A tinkertoy design based on stacked protection and deprotection schemes **24**
10.1 A preliminary: Some peptide chemistry 25
10.2 Stacked protection scheme and primitive operations 25

11 A tinkertoy design based on DNA hybridization **28**

12 An extremely simple planar tinkertoy design based on directed trees **28**

13 A highly rigid planar tinkertoy design based on directed trees **29**

14 Cleavage reactions **32**

15 Applications **33**

16 "Eutactic Nanosystems"? **34**

17 Conclusion **36**
17.1 5-year plan for future development of this area 36

18 Abbreviations used in this paper **38**

19 Acknowledgements **38**

20 References **38**

1 The problem

Suppose you would like to build a macromolecule shaped like the statue of liberty – or any other shape you care to specify. I have in mind the problem of synthesizing parts for "nanomachines" or "nanocomputers" using chemical techniques.

Thousands of synthetic methods have been developed over the years by chemists. For an introduction, see the books by Carey & Sundberg, Mackie et al., March (part 2 of which explicitly discusses 580 reactions), and Morrison & Boyd. All the reactions we use in this paper are described in these sources, especially March, unless we say otherwise. Larger collections are the 6-volume "Comprehensive organic chemistry," and the 9-volume "Comprehensive organic synthesis," (both Pergamon), "The chemistry of functional groups" (Wiley, >50 volumes so far), "Organic Reactions" (Wiley 1942-1997, 49 volumes so far), and L.Fieser & M.Fieser: "Reagents for organic synthesis," Wiley 1967 (with at least 16 additional volumes and a cumulative index meta volume now out). Other large multivolume treatises are discussed in March's appendix A.

But from my point of view (as a computer scientist) I would regard these as solutions to small sub-problems of the real, larger problem: we want an algorithm for the general problem of synthesizing *anything*.

Ultimately, you would like to be able to draw a chemical formula on your computer screen, and have it figure out how to synthesize it, and activate various robot arms which would pour test tubes, distill solvents, and so on, at the end of the day presenting you with a flask of the stuff you wanted.

Unfortunately, in its full generality this seems to be an extremely difficult problem – arguably NP-hard or even undecidable (Smith 1997).

So in this paper I'll pose an easier problem: that of devising an infinite subset C of all conceivable chemicals, rich enough to include "every possible useful shape," and devising a very small finite subset S of all possible starting materials and R of reactions, such that: there is a simple algorithm for synthesizing any chemical in C from S using R with approximately as high yield as one could possibly hope for, and approximately as simple a reaction sequence as one could possibly hope for.

I am interested in the structural properties – the shape – of molecules in C, and not particularly in their chemical or electronic properties, except insofar as they affect structural properties[1]. Thus I want every member of our class C to be structurally rigid and chemically stable.

2 The plan in brief

Our plan is to devise a small number of carefully designed "building block" molecules (or "tinkertoys"), which support a carefully chosen set of primitive operations which enable you to stick them together to create a fairly arbitrary subgraph of a spaceframe structure.

That is how to look at our plan from the "bottom up." The way to look at it from the "top down" is as follows.

To synthesize a molecule M shaped like the statue of liberty (to be specific, the intersection of the statue of liberty with a specific infinite periodic spaceframe structure) we split the statue of liberty using a bisecting plane, and then the problem is reduced to recursively synthesizing the top half M_1 and the bottom half M_2 of the statue of liberty, and then gluing them together. The entire recursively generated decomposition (rather like the "k-D tree" data structure of Friedman et al. 1977, which is easily generated automatically) is a binary tree. The leaves of this tree are single tinkertoys. The synthetic procedure starts with two tinkertoys which are leaves with a common parent, glues them together, and then at the next level this 2-component structure is glued to another 2-component structure to create a 4-component structure, and so on[2]. It is essential that at every stage of the synthesis, each of the two subtrees we plan to join to form a larger subtree can join in a topologically and geometrically unique way. To ensure geometric control we need every subtree that arises during the synthesis to be geometrically rigid (§8.1).

After M (but with only a subgraph – specifically a spanning tree[3] – of the desired bonds, call it M') has been constructed, we want to convert M' to M, thus changing the subset to the full set of all bonds and thus converting a somewhat rigid tree shaped molecule to a much more rigid, spaceframe molecule. This will be done by the application of a final "rigidizing reagent."

The fact that the splits are along these bisecting planes, forces the tree to be "assembleable," i.e. there is no steric hindrance to make it difficult for the two halves to approach each other so that the gluing reaction can take place. Any balanced tree-like decomposition forces only logarithmic depth. This log-depth property is important because our yields each reaction will be less than 100%. If 70%, then a sequence of K reactions will have final yield $.7^K$, i.e. exponentially small in K. Hence K had better be kept logarithmically small, which will assure final yields which only decline like a power law with the size of the final molecule[4].

Optimality statement. Assume all starting materials are of bounded molecular weight and each "reactive step" only creates one (or a small constant number) of new bonds and has bounded yield. Then logarithmic depth and yields which decline like a power law with the size of the final molecule, are both best possible.

Proof. Log-depth being best possible is a consequence of the fact that the size of the molecule can at most double each reactive step[5]; the optimality of power law yield immediately follows. □

---

[1] Actually, chemical and electronic properties might be very important for possible future developments, and might be highly controllable by the use of several kinds of tinkertoy components with different chemical and electronic properties. That is one reason Nature uses 20 kinds of amino acids; similarly we might imagine 20 kinds of tinkertoys. However, such issues will be left for future authors to resolve.

[2] Actually, things are not quite this simple (§6). The present discussion is intentionally simplified. We are trying to make it plausible that most useful shapes can be assembled in a treelike manner with only a logarithmic number of assembly stages.

[3] Note, there are *two* kinds of trees in this plan: the spanning tree M' of the molecular structure M we are synthesizing, and the recursive, hence treelike, organization of the synthesis itself. Don't confuse them.

[4] A molecule of size N built by joining operations starting from molecules of size 1 via a balanced tree of depth $\log_2 N$, where each joining reaction has yield y, will require a total mass of starting materials $N^{1-\log_2 y}$, i.e. growing polynomially with N. On the other hand, if the tree of joinings is maximally unbalanced (i.e. just a path of length $N-1$ with arms of length 1 hanging off it) then the total mass of starting materials required will be $\frac{2-y-y^{N-1}}{1-y}(\frac{1}{y})^{N-1}$, i.e. growing exponentially with N.

[5] Although the number of *steps* is only logarithmic, the total reaction *time* will not be. This is because the later steps, when one is joining larger pieces, will require more time than the earlier steps which join smaller pieces, due to larger expected waits before random diffusion lines everything up correctly. Ultimately, the total reaction time required should become *linearly* proportional to the number of atoms in the goal molecule. This could effectively set a limit on how large one can go. However, since rate constants vary over extremely wide ranges depending on the reaction and conditions (e.g. compare explosion of nitroglycerin vs. conversion of diamond into graphite) it may be that that the constant factor in the linear proportionality (by appropriate choice of joining reaction and conditions) may be made so small that this will be irrelevant. Or the bottleneck steps may not be joining reactions at all, which would also render this

3 Plan of the rest of the paper

Some chemists's immediate reaction to my ideas has been something like "that's already been thought of... protein folding, tectonics, self assembly, etc." In fact, although the ideas represented by these buzzwords are interesting, they are *not* at all the same idea as, and in fact fall far short of, tinkertoys. For why, and what the differences are, see §4.

We then reach the core of the paper: The building blocks, the spaceframe, and suitable sets of primitive operations on the building blocks to permit assembly.

We will first discuss (§5) the sort of properties that our molecular tinkertoys, and the spaceframe structure we wish to build out of them, must have.

Second, we discuss (§6) the mathematical question of what "assembleability" means.

Third, and centrally (§7): what abstract primitive synthetic operations should the tinkertoys support so that we can provably synthesize a very large class of structures?

Finally the sections at the end get less abstract and more chemical and physical. We survey some useful chemicals, reactions, and structural facts and finally lay out (§10-13) some reasonably complete specific building block designs which may actually work.

However, few chemical reactions can be planned entirely on paper (as I've done) and then trusted to work, so obviously a great amount of experimental work will be required – as well as additional theoretical work, since the plans I present, even if they do work, are surely improvable in many ways. The point of these sections is to lay out enough chemical details and design sketches to convince you that the abstract operations of §7 actually are likely to be implementable, compatibly, in practice.

Finally in §15-17 we discuss the likely course that future research in this area will follow, and discuss possible and impossible applications.

4 Comparison with previous work

The features which distinguish this paper from previous work are:

1. We propose (several) sets of abstract assembly operations which we prove suffice to assemble a very large class of N-brick objects with at most $O(\log N)$ synthetic depth.

2. Our process for assembling arbitrary shapes can be automated, and comes with an optimality statement.

3. By giving design sketches, we make it highly plausible that all of our operation sets may be implemented chemically[6].

4.1 Proteins

Mother Nature's process for building molecules of arbitrary shape and function is the DNA transcription and ribosomal process for creating arbitrary protein sequences from amino acid building blocks (Creighton 1993). However, our plan is a significant improvement on M.Nature's for several reasons. First, our synthesis proceeds in a treelike manner rather than a linear sequential manner, allowing our yields to decline only as a power law, not exponentially, and allowing our number of synthetic steps to grow only logarithmically, not linearly, as a function of the size of the goal molecule. (Admittedly, the amazing specificity and speed of enzymatic processes has allowed M.Nature to get very good constant factors, much better than the yields one presently can obtain in most synthetic procedures in organic chemistry.) Second, we directly synthesize a rigid, 3-dimensional structure via a simple algorithm, whereas M.Nature creates a highly flexible 1-dimensional structure. Although it is true that most natural proteins happen to self-fold rapidly into a well-defined 3D shape, only an extremely small percentage of possible protein sequences have that property, and their final structure usually still isn't very rigid. And the transformations mapping protein sequences to shapes or vice versa are algorithmically very hard to understand.

One may ask: If our plan is so superior to M.Nature's, then why doesn't Nature use it? The answer is that during the early evolution that created the biochemical processes found in simple cells, it was impossible to utilize our plan because it depends on the extra flexibility the chemist has by keeping different subtree chemicals in different jars, and choosing when to combine two jars to synthesize a larger subtree chemical. In primitive cells, there were no "two jars," there was only one jar – the cell itself. (Also, of course, Nature's whole plan for DNA and its replication would seem to make it much easier to stay sequential rather than treelike.)

---

irrelevant. This will have to be determined experimentally.

[6] We intentionally omit details about purifications and we also do not give syntheses of the tinkertoy designs we propose (instead we merely assume we have such tinkertoys available and then try to make it plausible that they will work).

A.Schweitzer has speculated[7] to me, however, that the present paper's plan may have been anticipated by Nature as "gangliosides," which are tree-shaped molecules, made of sugars, lipids, and modified amino acids, that may play a role in cell-cell recognition in multicellular animals. Gangliosides are synthesized in the "Golgi apparatus," a little understood part of the cell which contains many moveable vesicles which conceivably are able to serve as the "separate jars."

The story of how B.Merrifield was awarded the Nobel prize for building a "protein synthesis machine" based on repeatedly adding one more amino to a polypeptide chain anchored to a solid support, is recounted in the book by Mackie et al. Also mentioned there is the fact that by replacing the completely linear Merrifield synthesis by a partly linear and partly tree structured synthesis (see §10) the yield of the 58-amino peptide "Basic Trypsin Inhibitor" ($\approx$ 3500 Daltons) was increased by a factor of ≈ 3. Since we are trying to imagine syntheses of much larger molecules than this ($>$ 1000 building blocks?) we expect enormously better ($> 10^6\times$) relative yields for treelike syntheses versus linear ones. Also, since Merrifield's synthesis of the 124-amino peptide "ribonuclease" required 6 weeks, the savings in synthesis time achieved by switching to a treelike plan will probably also be essential.

4.2 Dendrimers

In some ways our plan resembles "dendrimers," which are tree shaped macromolecules created by organic chemists (see the review articles by Mekelburger et al. 1992, Newkome et al. 1992, Tomalia et al. 1990, and in Lehn 1996) which look, in micrographs, rather like fuzzy balls $100\mathring{A}$ across.

Linear, space-network, and treelike polymers were all synthesized at least 60 years ago and underly the plastics industry. For example, a tabletop made of "Formica" phenol-formaldehyde plastic is just one giant macromolecule consisting of chains linked together in a random appearing 3-dimensional network, any small part of which appears treelike. However, dendrimers go beyond plastics because they are *reproducible* macromolecules whose structural formula is *completely known.* (This has been verified by NMR spectra, mass spectrographic and other molecular weight measurements, and chromatographic means.) Conventional polymerizations always produce molecules whose masses vary uncontrollably within some range.

Of course, biological macromolecules such as hemoglobin also have reproducible and completely known structures and are the same approximate size as current dendrimers, but dendrimers have the advantage of being wholy synthetic. The prospect of making precisely identical fuzzy ball molecules, every "leaf" of which could contain essentially any desired chemical group (allowing controllable properties: for example high solubility, even for molecules this large, is achieveable by attaching solvent-philic groups to each leaf), has produced much speculation about possible great applications of dendrimers, but so far those hopes have been largely unrealized.

So far most dendrimers have been synthesized by a "divergent" process: one starts with the "root" of a tree, and then in stage 1 adjoins some number n_1 of "children" to that root, then purifies the resulting chemical. In stage 2, one adjoins n_2 children to each of the n_1 leaves of the tree from the stage 1 (these are now grandchildren of the root) and purifies that. And so on for some fairly small number, say s, of synthetic stages. The result is a tree shaped molecule with $n_1 n_2 n_3 ... n_s$ leaves, the simplest case being $n_0 = 3$, $n_1 = n_2 = ... = n_s = 2$.

As an example, here is the synthesis of Tomalia et al.'s Pamam dendrimer, which is particularly high yield and inexpensive. We start with a tree root (Pamam[0]) consisting of a single ammonia (NH_3) molecule. To get from Pamam[n] to Pamam[$n+1$], we first react it with CH_2 =CH-CO-OCH_3 (methyl acrylate) and then with $H_2NCH_2CH_2NH_2$ (ethylene diamine; a large excess is needed).

[7] And this is indeed highly speculative; although clearly many vesicles in cells are used for mutual isolation of two environments, and for transportation, at present I know of no evidence that any are used much as a chemist combines test tubes at will to control the course of a synthetic plan.

```
       CH =CH-CO-O-CH                           H N-CH -CH -NH
         2           3                           2    2   2   2
  NH  ----------------->  N(CH -CH -CO-O-CH )    ---------------->
    3                         2   2       3 3      in excess
 (Pamam[0])

        N ( CH -CH -CO-NH-CH -CH -NH  )    (Pamam[1])
              2   2         2   2   2  3

       CH =CH-CO-O-CH     H N-CH -CH -NH
         2           3     2    2   2   2
      ----------------->  ---------------->
                             in excess

         N ( CH -CH -CO-NH-CH -CH N ( CH -CH -CO-NH-CH -CH -NH )  )
               2   2         2   2      2   2         2   2   2 2  3

                        (Pamam[2])

       CH =CH-CO-O-CH     H N-CH -CH -NH
         2           3     2    2   2   2
      ----------------->  ----------------> Pamam[3] ...
                             in excess
```

In this way we extend the tree by 1 generation every 2-cycle reactive step, leading to a sequence of dendrimers

n	molecular mass of Pamam[n]	terminal amine groups	approx diameter (A)
0	17	1	2
1	359	3	10.8
2	1043	6	15.8
3	2411	12	22
4	5147	24	31
5	10619	48	40
6	21563	96	53
7	43451	192	67
8	87227	384	80
9	174779	768	92
10	349883	1536	105
11	700091	3072	Pamam[11] has not been synthesized.

In most dendrimers synthesized so far, the same chemicals have been used at every stage of the process, forcing one to stop at some fairly small value of s because the exponential "population explosion" in only polynomially growing available volume soon prevents further growth, at least if one wants that growth to be precisely reproducible. Thus, it seems to be impossible to synthesize Pamam[11], and no dendrimer I know of has mass exceeding 10^6 Daltons.

For some reason, none of the dendrimer researchers I know of have used "longer armed" chemicals at later generations of the divergent growth process, which would allow the population crisis to be arbitrarily *delayed*, allowing dendrimers with[8] significantly larger molecular masses. In my opinion it might be possible to near the 20th generation and reach record masses of over 10^9 Daltons, 10^6 leaves, and diameters of order 3000 A, the size of a small bacterium. *I think that it would be worthwhile to try this as a first step toward understanding the limitations that the tinkertoy proposal of this paper will encounter at record large molecular sizes.*

More recently, dendrimers have also been synthesized via "convergent" processes, starting from the leaves and moving toward the root. Convergent syntheses may sometimes be more immune to solubility problems because the large subtree reactants may be equipped with large numbers of solvent-philic leaf groups; they also might not require large reagent excesses to get high yields. On the other hand, the yields in convergent processes have suffered from steric hindrances which prevent the joining reactions.

As an example, here is Hawker & Frechet's (1990) convergent aromatic polyether dendrimer synthesis. We start with an initial one leaf tree "Hawker[0]-Br."

[8] The closest approach to this idea that I know of was Moore et al.'s "s.y.n.d.r.o.m.e." idea (see page 808 of the Dendrimer chapter in Lehn 1996), involving arms which became *shorter* as the generation number increased!

```
Ph-CH -O
     2  \
         C--CH
        //  \\
      HC      C-CH -Br                (Hawker[0]-Br)
        \    /    2
         C==CH
        /
Ph-CH -O
     2
```

To join two Hawker[n]'s to get a Hawker[$n+1$], we proceed as follows:

```
                 HO                             Hawker[n]
                   \                                    \
                    C--CH           K CO                 C--CH
                   //  \\            2  3               //  \\
2 Hawker[n]-Br +  HC      C-CH -OH  ------->         HC      C-CH -OH
                   \    /    2    18-crown-6          \    /    2
                    C==CH           ether                C==CH
                   /                                    /
                 HO                             Hawker[n]

          (3,5 dihydroxy benzyl alcohol)          (Hawker[n+1]-OH)

     ----------------------------->  Hawker[n+1]-Br
      triphenylphosphine and CBr
                                4
```

In this way Hawker & Frechet were able to synthesize Hawker[5]-Br, with mass 13542 Daltons, and then by reacting it with the special molecule CH_3-C(*para* Phenol)$_3$ to form a degree-3 tree root, they obtained a dendrimer of mass 40689 Daltons.

Some other dendrimer researchers have observed that it is possible to create *dipolar* dendrimers by, at the final stage of a convergent synthesis, adjoining two completely different subtree dendrimers to a common tree root. Also, H.Hart et al. 1986-1991 were able to make *rigid* hydrocarbon dendrimers, based on triptycene and in which the branching was all coplanar, with 12 leaves. (Although 12 sounds small, in fact it is the maximum possible, given their planarity and rigidity constraints. Other authors have made some other dendrimers of varying degrees of rigidity.) Our plan is based on pushing both of these two ideas to their ultimate limit.

Instead of always getting dendrimers that are complete trees and look like large fuzzy balls, we want to create any tree whatever with complete topological control, and also with complete geometric control because our molecules will be rigid.

4.3 "Tectonics"?

Finally, chemist T.Ebbesen told me that recently a movement in chemistry has begun, under the names "tectonics" and/or "modular chemistry," whose goal is to try to accomplish something similar to what I am proposing – precise control over the synthesis of macromolecular structures[9].

J-M.Lehn was awarded a 1/3 share of the 1987 Nobel prize for related investigations, and also was the editor in chief of the 11-volume "Comprehensive supramolecular chemistry" (1996) which contains a great deal of information relevant to the tectonics area. Moore 1995 is a short introduction to tectonics. The word "tectonics" was coined by J.D.Wuest.

More of the same will be found in Whitesides et al. 1991[10]. The idea is to synthesize medium size organic molecules which are designed (e.g. by placing positively and negatively charged groups) to crystallize in ways that will create large voids in each crystal cell. The hope is that such voidy crystals will have applications: for example, it may be that the chemical environment inside voids will encourage or discourage certain chemical reactions, so that the voidy crystal might be a useful catalyst for some (as yet unknown) chemical process. More generally, by such techniques one might hope to make medium size mostly-organic molecule building blocks which will magically

[9] This isn't very far advanced yet, because searching *Chemical Abstracts* under the subject heading "tectonics" only yields articles about geography and continental plates, and "modular chemistry" seems to be an educational method in which chemical notions are introduced to students in bite size pieces!

[10] In which, incidentally, synthetic approaches of the sort I discuss are dismissed in 1 paragraph!

self-assemble (due to non-covalent intramolecular forces) into interesting periodic, micellar, or other structures of moderate Kolmogorov complexity.

The most impressive example of this that I know of was the two part article Anelli et al. 1992 and Annabilino et al. 1995, in which, e.g. "*catenanes*" with up to 5 "chain links" self assembled and then self-bonded[11].

Some *viruses* are known to self assemble from their component macromolecules, again via intramolecular forces. *Perfect crystals* are of course also large self-assembling structures in which every atomic location is known. (Although in practice, crystals are usually imperfect. Tinkertoy constructs are like intricately carved perfect crystals.)

It seems to me, that such things, while interesting, are just groping in the dark, compared to the tinkertoy strategies outlined here. All self assembly techniques are like tossing 50000 bricks in the air and seeing what will result. If great care, artistry, and thought is devoted to the sculpting of each individual brick, then something moderately interesting may result, although surprises will occur[12]. But meanwhile, tinkertoys are more like a systematic procedure for using bricks to build houses, and neither care, artistry, nor thought are required. Instead we have a *general* synthesis *algorithm* for getting precisely what you want with no guesswork, and this procedure is automated. Also with tinkertoys, covalent bonds are used, rather than intramolecular forces, which are regarded mainly as a parasitic effect that should be minimized[13].

These differences may be quantified in terms of *Kolmogorov complexity.* N-tinkertoy constructs will achieve Kolmogorov complexity $\theta(N)$, while dendrimers, perfect crystals, and self-assembling micelles have Kolmogorov complexity $\theta(1)$. In short, dendrimers, micelles, perfect crystals (and probably self assembling structures generally) are boring, tinkertoy constructs aren't; "Kolmogorov complexity" is a notion from computer science that quantifies "boring."

4.4 The big and small pictures

Because in the present paper I am approaching the subject from the viewpoint of a computer scientist and mathematician, I've probably come up with some new kinds of thoughts, from the viewpoint of chemists. In particular, I'm trying to think about the "big picture" – strategy – before working out all the chemical details – tactics. This advantage of this approach is that one can lay out and compare complete research plans to achieve the desired goal, but the disadvantage is that ultimately the bulk of the work required to carry out such a plan *is* going to be in the chemical details and will require a very large amount of theoretical and experimental work.

5 Building blocks ("tinkertoys") and spaceframe

I'm imagining a building block molecule which looks rather like a hub with several rigid arms sticking out of it. At the ends of these arms are chemical groups ("hands") specific to that arm. The plan is to join tinkertoys at their hands with reagents that show specificity for the particular hands we want.

At first the simplest scheme seems to be to have only one kind of universal building block – and in principle, as we'll see, that will indeed suffice – but in practice it is might be easier (and certainly would increase the usefulness and flexibility of the tinkertoy set) if several different kinds are available with various combinations of different kinds of arms, including amputated arms, plus various optional decorations.

5.1 Only three arms are needed, and spaceframes that prove it.

The first question to ask is: how many arms do we need on each building block? Certainly we must have at least 3 arms, because it is not possible to make nontrivial tree structures all of whose nodes have degree ≤ 2. And we claim that 3 arms in fact suffice to construct interesting infinite periodic spaceframe structures.

[11] These authors also use the word "tinkertoys" but they do not mean what I do by it.

[12] For example, one problem that arose in one of Wuest's self assemblies (Moore 1995) was the unintended creation of two *interpenetrating* lattices. With only small care in the design of our tinkertoys and the structuring of our treelike assembly process, unintended interpenetration and catenation should be impossible. This is because by the time that "global rigidization" happens, creating cages which could in principle trap something inside them, nothing small enough to get trapped will be left. Also, every time we join two large subtrees, allowing the possibility of trapping something deeply inside their (roughly planar) interface, again nothing small enough to be trapped will be present.

[13] However, we do *not* wish to disparage self assembly and intramolecular forces completely. At the hypothetical "next level of the size heirarchy" (§7.3) such forces and strategies might usefully be employed. Clearly, Nature uses self assembly and intramolecular forces to make some interesting structures. For example the simple Tobacco Mosaic Virus (40M Daltons) self assembles from 2130 identical protein units, plus an RNA. The *E.Coli* ribosome (2.5M Daltons) self assembles from about 55 (nonidentical) proteins and 2 RNAs. My point is that, it appears from these and such artificial examples as Zhang & Seeman's (1994) "DNA truncated octahedron" (§7.2; 760K Daltons and made of 14 pieces) that in order to get effective results, one's pieces need to be molecules of completely controllable structure *with sizes in the* $50K$ *Dalton range.* One may also argue from first principles that such sizes are required because noncovalent interactions are $\approx 100\times$ weaker than covalent bonds, so one needs to glue pieces with > 100 precisely placed such interactions, which requires large pieces! It seems to me that the best, maybe the only, way to get into this size ballpark is to use the techniques of the present paper.

There is more than one way to see this. One method is as follows. By joining two planar 3-armed structures lying in perpendicular planes containing as their common line the line of the joining arm, we get a 4-armed structure whose 4 free hands form the vertices of a tetrahedron. Recall that the diamond crystal is constructed of 4-armed entities (tetrahedral carbon atoms). We similarly plan to construct a (stretch-distorted) diamond-like framework using our 4-armed structures. Note that there is no necessity to have exact regular tetrahedral symmetry – any tetrahedron, plus its mirror reflection, will do. The graph of the diamond crystal is bipartite: the "red" carbons are joined only to "blue" carbons and vice versa. Thus each red carbon is joined (as a second neighbor) to 12 red carbons, and these red carbons in fact form an FCC lattice. (The diamond crystal in thus a union of a red and a blue FCC point lattice.) The prescription we have devised using non-regular tetrahedrons (red) and their reflections (blue) will result in a distorted FCC lattice unit cell parallelipiped, but note that all the reflections "cancel out" so that the bonds which need to be parallel to the same bonds at the opposite side of the parallelipiped, in fact are exactly parallel if arms of the same type (only mirrored) are always joined.

Another way to look at this is: Two complementary chirality tetrahedrons join to form a 6-armed object which is its own mirror image. The 6 arms come in parallel pairs. These 6-armed objects can be used to tile space according to (a distorted version of) the usual 3D cube wireframe grid.

Another construction is this. Just using 2 of the 3 arms suffices (now assuming some turn angle at the joined hands) to construct a helix with the third unused arms pointing outward. These helices, assumed to have parallel axial lines arranged at the vertices of a 2D equilateral triangle lattice, may then be joined via the 3rd arms, resulting in a periodic 3D network.

If this helical scheme is to work, we need rather precise geometry - the 3rd arms must emanate at 60 degree angular intervals. (Or 120 degrees for the vertices of a 2D regular hexagon tiling, or 90 degrees for the square lattice.) The previous "distorted diamond" idea did not need careful geometry to get angles right due to the automatic "canceling" effect of the mirror reflected antichiral tetrahedron. This might seem to represent an advantage for the distorted diamond idea, except that when joining the two tetrahedrons to make the 6-armed object we need to be careful to make the twist angle at the join be exactly zero so that the final object is its own mirror image. Hence some precision is needed in the diamond idea too. Another disadvantage of the diamond idea is that we need to have two chiralities of building blocks available – unless the two are the same (probably the simplest thing to do) – which may not be easy.

Wells 1977, 1979 initiated the systematic study of "three dimensional nets," which are infinite graphs embedded in 3-dimensional space (with line segments for the edges and points for the nodes) in a manner symmetric under some 3D crystallographic group. Most interesting for us are Wells's "type A" (transitive on edges) and "type B" (transitive on points) nets[14]. The current knowledge of such nets is incomplete, but complete knowledge of all point sets transitive under a 3D crystallographic group is available (see the "International tables..." 1988) so one "only" has to consider the ways to link up the points.

Incidentally, the "bipartite graph" idea, which came up several times above (where there are two kinds of building blocks – red and blue – with the only joins being between blocks of differing colors) is very useful for simplifying many synthesis geometrical matching tasks. For example, if you are trying to create a spaceframe structure like the 3D cubic grid, 3D diamond structure, or square or regular hexagon tesselations in 2D (all of which are bipartite graphs) then you usually can design your nodes with intentionally imperfect angles because by using antichiral blue nodes and red nodes, you will get exactly opposite canceling angles resulting in a "2 nodes joined by an edge" substructure that, as far as the outside world is concerned, is geometrically perfect. (For an example, see the fused pyrrole links in §13.)

Finally, although 3 is the minimal number of arms required, in some scenarios it may be simpler to rely on a 4-armed tinkertoy, not a 3-armed one. For example, suppose we want to build 2-dimensional tinkertoys for use in building arbitrary *planar* structures. Then with a tinkertoy with 4 arms and square symmetry, only 2 kinds of fundamental joining reactions would be required (vertical kind and horizontal kind), not 3 as one might think would be needed for a planar 3-armed tinkertoy with equilateral triangular symmetry.

5.2 The need for porosity; the unimportance of size limits

The "wireframe structure" we plan to build needs to have comparatively long "wires," (i.e. our tinkertoys need to have long arms) so it forms a highly porous, low density network. This porosity is required both so that various reagents can penetrate the network to reach internal reactive sites, and also so that two large subtree molecules can approach each other to join together into a larger tree, without having solvation surface energy effects preventing that approach. Ultimately for large structures, the accumulation of many small (non covalent bond) forces will energetically outweigh any one covalent bond. For that reason, for our treelike synthesis strategy to work, it ultimately will be essential to keep these extraneous forces small. In principle this ought to be possible by simply

[14] Treacy 1996 has done further work in this area by computer search.

keeping the structure very porous (in the limit of infinite porosity, of course solvation surface energy effects are negligible, but we can't achieve enormous porosity without sacrificing rigidity) and or by attaching solvent-philic or phobic groups to each tinkertoy. The plan is to "null out" solvation surface energy effects by adjusting these "knobs."

For very large macromolecules, just the Brownian *fluctuations* in the forces on each half will become sufficent to rip the two halves apart, if the only thing holding them together is a constant number of covalent bonds. This sets an upper limit on the size of the largest molecule that can be assembled by a treelike assembly strategy. However this limit is extremely large. The fact that it is large has already been seen by considering artificial dendrimers (§4.2), but the fact that it is extremely large may be seen as follows:

1. Each of our chromosomes is a linear DNA molecule averaging over 10^8 base pairs (mass 7×10^{11} Daltons) whose halves are held together by only two bonds;
2. The Congo eel *Amphiuma* has chromosomes about 40 times larger (and some plant chromosomes are still larger);
3. With great care to avoid viscous shearing, intact DNA strands up to 4cm long have even been prepared and radioautographed in fully *unwound* form.
4. Block et al. 1990-1997 have done several experiments in which single molecular strands were used to attach micron sized beads to surfaces.

Eventually after the application of the "rigidizing reagent" our structures will become entirely robust since there will no longer be any battle between many small forces and one covalent bond, instead it will be many small forces versus *many* bonds.

6 Assembleability

In this section we will define and investigate the mathematical notion of "assembleable" shapes. We will see that assembling shapes in the worst case is very difficult (except for planar shapes with motions being allowed to be made in 3-space, which are easy), so then in §6.1 we resort to the time honored trick of defining away the problem, by defining the class of "useful" shapes (ones which are easy to assemble) and then demanding them. Arguments are made that the "useful" shapes encompass most or all of the shapes people actually want, and in §6.2 we define and show how to find the best possible "iso-oriented tree sliced" assembly of any shape.

Definition: A (not necessarily convex or simply connected, but necessarily connected and bounded) polygon in the plane will be said to be "assembleable" if either

1. it is convex, or
2. if it may be subdivided into exactly 2 assembleable polygons by a piecewise linear curve consisting of at most some constant number of line segments, such that these two parts may be brought together by some continuous rigid motion starting from infinity in such a way that at every point in time, the two parts are interior-disjoint, and at the conclusion of the motion, the original polygon has been regenerated.

Definition: A polygon (as above) is "assembleable in N gluing steps with synthetic depth D" if the binary tree of recursions generated by the previous recursive definition (the leaves of the tree are convex polygons, the interior nodes are gluing operations, and the root node is the completed, fully assembled polygon) has N interior nodes and maximal depth D.

Similar definitions pertain to 3-dimensional polyhedrons in 3-space (the cut which is a piecewise linear curve becomes a cut which is a piecewise linear surface made of at most a constant number of triangular "panels") or to 2-dimensional polygons but with the motions being permitted to occur in 3-space.

Theorem: All n-gons in the plane are assembleable, if motions are permitted to occur in 3-space, in $\leq n-4$ gluing steps and with synthetic depth $\leq \log_{3/2}(n-2)$.

Proof sketch: Any n-gon may be divided by diagonals into $n-2$ triangles. The adjacency relations among these triangles form a tree whose vertices (triangles) all have degree≤ 3. By finding a "splitting edge" (Kang & Ault 1975) in this tree we may find a diagonal which splits the n-gon into an a-gon and a b-gon with $a+b=n+2$ and $\max(a,b) \leq 2n/3+2$, and this will be the edge we use to divide the polygon into two parts. There is never any problem finding suitable continuous motions to bring the two parts together, if we have access to the third dimension. □

Theorem: There is no function $F(n)$ such that all n-gons in the plane are assembleable with $\leq F(n)$ gluing steps. (And a similar negative result holds for polyhedrons in 3-space.)

Proof sketch: The counterexample polygon will be a 2-armed jagged spiral whose two arms are joined at the outside. The arms have very small width ϵ^2 at all windings of the spiral except at the innermost, where the width is of order 1. The moat between the arms (part of the exterior of the polygon) has width ϵ everywhere. The total area and diameter of the polygon will both be of order 1.

Now, if this polygon is to be divided into two parts by a piecewise linear curve as in the definition of "assembleable" above, then we can see that one of the parts must have area $\leq 2.9\epsilon^2$. This is because the small part is attached somewhere inside of the spiral, it has to be moved in through the channel of width ϵ, and the requirement that it fit through corners in the channel forces its length to be $\leq \sqrt{8}\epsilon$, so its total area is $\leq \sqrt{8}\epsilon^2$. On the other hand, if the small part is attached near the outside of the spiral, it has to be of width $< \epsilon^2$ to avoid cutting the big part in half, and its length must be ≤ 1. (You cannot cut into two big pieces because they are entangled and inseparable by any continuous rigid motion.) Now once the small part is removed, the same argument continues to apply (perhaps with slightly larger constants) for at least order ϵ^{-2} "bites." Notice that here ϵ may be made arbitrarily small completely independently of the value of n, provided n is sufficiently large. $\square$

On the other hand:

Theorem: Every polygon in the plane and every polyhedron in 3-space is assembleable. Furthermore, for polygons (polyhedrons) made by gluing together $N+1$ unit squares (cubes) at common faces, the assembly can be done in N gluing steps; although even then synthetic depth $D \geq 0.09N + O(1)$ can be forced by suitable spiral polygons.

Proof: This is seen by dividing the polygon (polyhedron) into a sufficiently large number of sufficiently small triangular (tetrahedral) elements. If the elements are small enough (smaller than the minimum "gap size"), they will always manage to get past any obstructions. The statement about unit squares and cubes follows similarly since the gap size is 1 and always glue on a cube which is a leaf of a spanning tree of the adjacency relations among all the cubes. Just using a double spiral joined at the outside made of N squares in a $\approx \sqrt{2N} \times \sqrt{2N}$ grid will suffice to force synthetic depth $D \geq 0.09N + O(1)$ by a similar argument to the previous spiral argument. $\square$

Conjecture: A random spanning tree of the square (cubic) wireframe $N \times N$ $(\times N)$ grid will require (with high probability when N is large) synthetic depth that grows faster than $N^{1-\epsilon}$ for any positive ϵ.

This conjecture would indicate that random complex structures cannot be synthesized in high yield.

Conjecture: The problem of determining the minimum number of gluing steps, or the minimal synthetic depth, for assembling a polygon in the plane (or a polyhedron in 3-space) is NP-hard.

There are really two separate parts to the problem in this conjecture. First there is the question of the minimum number of parts needed in a subdivision of the polyhedron. Second, there is the question of how to maneuver those parts to perform the assembly. Quite a lot is known about problems of both the first and the second type.

Chazelle 1980 (see also Chazelle & Dobkin 1985) showed that it is possible to determine the minimal number of convex pieces one can decompose an N-gon into (and find such a decomposition) in only $O(N^3)$ computational time. However, for polgons with polygonal holes, or for 3D polyhedra, the corresponding problem is NP-complete. These efforts combined with those of Lipski et al., Chazelle & Palios, Lichtenstein, Lingas, O'Rourke & Supowit, and Ruppert & Seidel (see the references and the book by O'Rourke 1987; also I have put a few easy results in the table which may not be mentioned in these sources), managed to establish the results in table 1.

The question of how best to assemble (or even of whether one *can* assemble) a polygon or polyhedron out of specified convex pieces (where the final location of each piece is known) in a treelike manner (i.e. always joining only 2 pieces at a time) has not been studied. But various similar sounding problems are known to be NP-hard and sometimes even PSPACE-hard. For example, you may have seen the ancient Norse toy (called the "Chinese rings" in Berlekamp et al. 1982 pages 750-753) in which a long U-shaped loop can only be un-entangled from N rings attached to sliding widgets after making 2^N elementary motions. Such questions of separability by simultaneous motions of N entangled rigid polyhedral objects are definitely at least PSPACE hard because a computer may be built out of such objects and the separability question made equivalent to the question of whether some input exists that will cause some PSPACE program to halt. (And see Chazelle, Ottman et al. 1984 for a proof that the same problem for rectilinear polygons moving in the plane, is NP-hard, and also may require an exponential number of elementary moves.) On the other hand, the question of whether *one* polyhedron may be moved from one specified placement to another despite a finite number of polyhedral obstacles, is solvable in only polynomial time, because in the 6-dimensional space of locations and orientations of the moving object, one can build a polynomial-size diagram describing all the legal ones. Then it is a simple matter of seeing whether the start and goal points are in the same connected component of the diagram and finding a legal path between them if so. The same approach will work to find a simultaneous motion of N polyhedra to get from a start to a goal state (or prove that impossible) but since the diagram will be in a (much huger) $6N$-dimensional space, this algorithm will be incredibly inefficient, consuming both exponential space and time. But J.Canny has shown how to solve this problem while consuming only polynomial space (e.g. see Renegar 1989), so it is in fact PSPACE-complete. Thse facts are summarized in table 2.

Object being decomposed	Decompose into	Holes?	Computational complexity
rectilinear polygon	min # rectangles	rectilinear polygons	$O(N^3)$
polygon	min # convex polygons	no	$O(N^3)$
polygon	min # triangles by diagonals	no	$O(N^3)$
rectangle	min # rectangles	points	NP-complete
polygon	min # convex polygons	polygons	NP-complete
polyhedron	min # convex polyhedra	don't care	NP-complete
polygon	min # triangles by diagonals	polygons	NP-complete
polygon	min # triangles	polygons	NP-complete
polygon	min # convex polygons by diagonals	no	$O(N^3)$
polygon	min # convex polygons by diagonals	polygons	NP-complete
polyhedron	tetrahedra, min # "Steiner points"	don't care	NP-complete
star shaped polyhedron	tetrahedra by diagonals	no	NP-complete
polyhedron	tetrahedra	don't care	$\Omega(N^2)$, $O(N^2 \log N)$
polyhedron	convex polyhedra	don't care	$\Omega(N^2)$, $O(N^3)$
polygon	triangles by diagonals	no	$\theta(N)$
polygon	triangles by diagonals	polygons	$\theta(N \log N)$

Table 1: Worst case computational complexity of various geometric partitioning problems. N is the number of edges of the polygon or polyhedron and its holes, i.e. the size of the input.

Motion problem	Computational complexity
Simultaneous 3D motion of N rigid objects	PSPACE complete
Simultaneous 2D motion of N rectilinear polygons	NP-hard, can take exponential # moves
3D motion of 1 object among N rigid obstacles	Polynomial time

Table 2: Worst case computational complexity of various motion planning problems

6.1 "Useful" shapes

The negative results and conjectures above seem very unfortunate, but for most of the *simple, nonrandom* shapes one is going to be interested in synthesizing, assembly is not so hard.

Definition: The "useful" polygons and polyhedra are the ones which are assembleable in N gluing steps with synthetic depth $O(\log N)$, and such that it is easy[15] to find a recursive cut sequence that proves it. Here N is less than or equal to the reciprocal of the volume of a ball whose diameter is the smallest feature size (e.g. width) in the object. (I.e., N is the number of "pixels" in the object.)

I suspect that most or all of the the polygonal and polyhedral shapes that arise in engineering are useful. In fact, for a large class of objects, just making "greedy" planar knife cuts parallel to the coordinate planes at each stage (e.g. that chop off the biggest possible smaller chunk) works. For example, consider a large office building honeycombed with rooms, corridors, pipes, and so forth. Such shapes are certainly in our class and should be synthesizable with a process very much like the greedy or k-D tree algorithm.

In the rest of this paper, we are only going to be interested in useful polygons and polyhedra – unless we are considering the case of plane polygons with 3D assembly, in which case, as we've shown, *every* polygon is useful.

6.2 Polynomial time algorithm to find the best "iso-oriented tree sliced assembly" of a lattice animal

Definition: A "lattice animal" is a connected subgraph of an infinite periodic graph such as the square or hexagon tessellation in the plane, or the cubical grid in 3-space.

Definition: A "tree assembly" of a lattice animal is a way to construct the lattice animal starting with single lattice nodes and gluing them together with rigid rods to construct spanning trees of subgraphs of the lattice animal, which in turn are glued to make spanning trees of still larger subgraphs, until eventually we have a spanning tree of the whole lattice animal.

Definition: A "tree sliced assembly" is one in which every time two trees are glued together, the two sub-animals represented by the two trees are separable by a hyperplane. If the hyperplanes's orientations always are selected from a fixed finite set (e.g. lines parallel to the x and y axes in the plane) then we have an "iso-oriented tree sliced assembly."

[15] You can use whatever definition of "easy" you like, but one possibility is just to use the algorithm of §6.2. That way we will have an explicit algorithm to decide whether a shape is "useful."

Theorem: Fix the periodic supergraph. Assume some polynomial-time subroutine is available[16] to compute the "cost" of a tree assembly from the subtrees of its root node and their costs. Then there is an algorithm whose time and space consumption is bounded by a polynomial in N, which will find the least-cost iso-oriented tree sliced assembly (or if no such assembly exists, will say so) of any N-node lattice animal.

Proof. We use "dynamic programming." The idea is that there are only a polynomially large number of possible subsets of the nodes of our lattice animal which can be got by using iso-oriented hyperplane cuts[17]. We make a table of all possible such sub-animals and compute the least cost assemblies of each, starting from 1-node animals (cost= 1) and computing the costs of larger-area animals from the already-computed costs of all their subanimals by considering every possible way to slice it into two. By also storing pointers to the two subanimals in the table, the assembly, as well as just its cost, is computed. □

Definition: A "rooted lattice animal" is a lattice animal one of whose boundary nodes is denoted "root."

Definition: A "rooted tree sliced assembly" of a rooted lattice animal is one in which all sub-animals arising in the assembly are rooted spanning trees and any time two sub-animals are glued together, it is by gluing their two root nodes together and designating one of these nodes as the new root of the union animal.

Theorem: There is an algorithm whose time and space consumption is bounded by a polynomial in N, which will find the least-cost iso-oriented rooted tree sliced assembly of any rooted N-node lattice animal.

Proof. Same proof as above, but the table is now of all possible (subset, root-node) pairs. □

Remark. A simple algorithm for finding a usually low-cost, but not necessarily sliced nor optimal, rooted tree assembly of a lattice animal L, is to start with two nodes A and B at largest possible graphical distance apart, and then glue A to its neighbor A' on the shortest path from A to B. The same operation is then repeated on $L\backslash A$.

7 Sets of Primitive operations

7.1 A set of primitives with universal assembling power based on stacked protection – only one kind of building block needed

.

Definition: A "primitive operation" is some sequence of chemical steps that transforms every instance of some particular chemical group occurring in all the molecules in a jar, in some particular way.

For the following set $\{P_{kn}, D, J\}$ of operations, only one kind of building block is needed, with k arms ($k = 3$ or $k = 4$).

Operation P_{kn} [initial stacked protection]: Adjoin a stack of n protective groups to arms of type k.

Operation D [de-protect]: Remove 1 level of protection from every protected arm, thus converting each arm with stacked protection $n > 0$ levels deep, to an arm with $n - 1$ levels of protection.

Operation J [join]: Given two subtree molecules with exactly one unprotected arm each, join them at their hands. (Note, several different J variants might be required, because there are several kinds of hands. But actually, in principle there only needs to be *one* kind of hand, and thus only one kind of joining operation, provided the different hands on each building block somehow had been provided with different and appropriate levels of protection ala P_{kn} before beginning the synthesis. Also, even with k kinds of hands, it is not necessary to have all k^2 variants, for example for the assembly of planar molecules from square 4-armed building blocks, in fact only 2 variants, not 16, are needed. Also, some joining variants may function even in the presence of certain unprotected hand groups, so long as they are of incompatible types, which can simplify syntheses.)

Theorem: Operations P_{kn}, D, J plus unlimited supplies of the k-armed building block molecule, $k \geq 3$, will suffice to construct any tree graph structure with nodes of degree $\leq k$. It suffices if P_{kn} are only applied to jars containing exactly one building block, i.e. initially, and J and D operations only are done thereafter. All synthetic pathways need only have logarithmic length.

Proof: Proof is by induction. At each stage of the synthesis we deprotect the trees from the previous stage by using D, resulting in trees with exactly one unprotected group (because all the n-values during the initial n-level protections P_{kn} of single building blocks [the tree leaves] at the basis of the induction were cleverly chosen), then combine the jars containing two subtrees and join them with J. That proves we can synthesize any tree. To prove we can always do so with only logarithmic depth, we use the fact that any n-node tree has a node called a "centroid"

[16]For example, if gluing two subtrees with a new rigid rod is a chemical reaction with estimated yield y, $0 < y < 1$, then one might take as the cost of the resulting tree, $1/y$ times the sum of the costs of the subtrees, plus R, where R is the cost of the reagents and the time required to carry out that reaction. The cost of a single node would be 1.

[17]For example, if the lattice animal is an N-node connected subgraph of the planar square grid, and the permitted orientations are parallel to the grid directions, then the only possible sub-animals are rectangles and there are $\leq (N + 2\sqrt{N} + 1)N/4$ of them.

which can be found by a simple algorithm in $O(n)$ time (Kang & Ault 1975)[18]. Removal of the centroid of a tree will cut it into subtrees, each of which has $\leq n/2$ nodes. These subtrees may similarly be decomposed, the entire resulting decomposition being $\leq \log_2 n$ levels deep. We synthesize the subtrees recursively and join them to a new centroid node. Note that the initial protection chain lengths all only need to be $O(\log n)$ long. □

Although this set of operations suffices for creating any *graph theoretic* tree, in practice *geometric* difficulties may arise. Of course, the negative results on assembleability in §6 show that our logarithmic depth result is in general impossible to achieve, but if we assume as usual that the desired shape is such that assembleability is easy, that is no problem – but we may *still* have problems because the long protective chains may serve as steric hindrances for the joining reactions. This is avoidable by making the arms of the building blocks much longer than the protective chains. What might be a better avoidance is to add to our repertoire the operation

Operation K_k [kill]: Kills the reactivity of all unprotected arms of type k.

and/or (what is almost the same thing) we could provide additional types of building block starting materials in which some of the arms are missing. The idea is to take advantage of the fact that most of the free hands are never going to be joined to anything, so why bother protecting and deprotecting them when you could just kill them.

If we also add

Operation U [un-kill]: Restores reactivity of all killed arms.

Operation G [global rigidize]: Joins every pair of nearby and compatible un-killed and un-protected hands together[19].

then we become capable of making lattice-like structures, not just their spanning trees.

7.2 Alternative set of universal assembling primitives based on building blocks with DNA arms

The plan of §7.1 above used a unique (or only a few kinds of) building block and achieved the chemical specificity we needed during joining reactions solely by use of multi-level protection P_{kn} and de-protection D operations.

At the other extreme, we could in principle abandon protection and instead get specificity by the use of a very large number of different kinds of hands on our building blocks.

Consider building blocks with 1-4 "hands" consisting of short specific single stranded DNA sequences. (Actually double stranded DNA will be wanted for reasons of angular rigidity, see below, but one of the strands will be a lot longer than the other, leading to a long single stranded "sticky end.") Upon mixing, two building blocks hopefully will join hands if and only if those hands have complementary and reversed DNA sequences. Again we mix subtrees in pairs, again at each stage exactly one joining reaction occurs, and again we may use the tree-centroid decomposition to assure logarithmic synthetic depth.

Although the resulting creations will only be held together by (weak) hydrogen bonding between the complementary bases in the "sticky ends," we may then "ligate" the DNAs to create covalent bonds attaching the joined hands permanently and leaving no "loose ends" anywhere. Although the usual DNA-ligation method depends on a "ligase" (enzyme) which is probably too large to fit inside the macromolecule we are assembling, recently high-yield non-enzymatic techniques have been developed for DNA ligation (Herrlein et al. 1995).

Theorem: The two operations of (1) initially adjoining DNA double strands with an $O(\log N)$ long "sticky end" with specified sequences to each arm of a k-armed building block molecule, and (2) of combining two jars so that their strands hybridize if they are complementary, suffices to synthesize any tree with N nodes, each of degree $\leq k$, from tinkertoys with k DNA arms each with sticky ends of length $O(\log N)$. This is assuming mismatching hybridizations are forbidden. □

Advantages: This DNA-based scheme seems simpler than our previous scheme based on "stacked protection," and it takes advantage of the present state of high technological development of DNA manipulation and synthesis.

Disadvantages: DNA-based molecules are not very stable chemically, nor are they very rigid; their persistence length (§8.1) is only equivalent to ≈ 150 base pairs, i.e. only ≈ 10 tinkertoy arm lengths. Also, the whole scheme depends on the lack of mismatch hybridizations. Many kinds of mismatch pathologies (e.g. "hairpins" and self matches) are known. Of course, mismatches are discouraged naturally because they incur an energy penalty roughly equivalent to 1-2 "hydrogen bonds," i.e. ≈ 8 Kcal per mole of mismatches, but this penalty is much smaller than a *real* bond, of energy ≈ 100 Kcal/mole, which is what one got with the previous non-DNA scheme.

These disadvantages may be partially overcome as follows.

First, "**Operation** G" (global rigidizing) may again be possible in this picture as follows:

Make the "killed" hands of the arms be DNA with especially short (so they don't get in the way during tree assembly) sticky ends with the same sequence always, e.g. CCC. Then to form links between all CCC arm-stubs at

[18]Centroids imply a recursive decomposition of any n-node tree which is $\leq \log_2 n$ levels deep. In fact, the *best* such decomposition (which is at least as good) may be found efficiently, for a large number of meanings of "best," by "dynamic programming." The idea is that an n-node tree has only $2n-2$ subtrees, and a table giving the cost of each may be constructed, small subtrees first.

[19]Note, all the tinkertoys were already rigidly fixed in their perfect lattice locations and orientations, by the rigid spanning tree. The global rigidizing process just cements them there.

the right distance apart, introduce double stranded DNA with 3-long "sticky ends" GGG on each side. Hopefully this DNA will pair on both sides, although, statistically speaking, this seems less likely than just requiring one end to pair as before.

Also DNA strands may be "protected" during such operations by hybridization with free complementary strands, and then deprotected by (e.g. thermal) denaturization. Denaturing conditions will not destroy linked arms because they were *covalently* ligated.

By the use of global rigidizing operations during intermediate stages in the synthetic process, as well as at the end, we may be able to overcome the rigidity problem.

Second, by clever choice of the DNA sequences we use, it should be possible to discourage mismatches more than with a poor choice. For the combinatorial problem (which is rather like the problem of constructing "error correcting codes," but with a rather peculiar "distance" metric) of constructing sets of short DNA sequences with properties that try to avoid unwanted inexact pairings and hairpin self pairings, see the appendix of Smith 1996 on "modified DeBruijn sequences." For elementary background knowledge about the chemistry of DNA, see Saenger 1984, for example.

There has been quite a bit of previous work in this area.

First, Nadrian Seeman and collaborators (Chen & Seeman 1991; Li, Yang et al. 1996) had previously used DNA and "Holliday junctions" to form such shapes as a "wireframe cube" made of 6 intertwined and topologically linked closed loops of DNA. (Each DNA single strand constitutes one square face of the cube. Each edge of the cube is double stranded DNA formed from hybridization of the single strand DNAs circumnavigating each neighboring face.) Later they made a similar wireframe "truncated octahedron" (Zhang & Seeman 1994; this has 36 edges and 24 vertices and masses 7.9×10^5 Daltons) and various knots and catenanes (Seeman et al. 1993). The ultimate yield in the truncated octahedron construction, starting from its individual DNA sequences, was apparently about .000004. The Seeman group's motivations, methodology, and the sorts of chemicals and structures they produced were different from the suggestions we are making here. In particular, we do not need or want Holliday junctions or catenation. Also Seeman was mainly driven by the different and difficult ultimate goal of making "artificial crystals" to enable X-ray crystallography to be used on substances which normally cannot be crystallized. Nevertheless, some of the techniques they developed should be useful in further advancing our own suggestions.

Specifically, the ideas they required for the truncated octahedron were as follows. "Solid support" methodology was employed, and thanks to "ligations," every intermediate synthetic product consisted entirely of topologically linked closed loops. Therefore, purification at every stage could be accomplished by digesting all DNA strands which had "loose ends" by an appropriate nuclease, followed by simple washing[20]. Now for our DNA-based scheme (although it does not use Holliday junctions nor closed loop DNA) loose ends also get eliminated by ligations[21]. Therefore, the same digestion-based purification idea might still be possible in our DNA scheme. The "solid support" idea is also compatible with all of our schemes (including the non-DNA ones), except that in our schemes, half of our reactants would have to be detached from their respective solid supports before each new joining step.

Also, A.Schweitzer pointed out to me the suggestively titled paper Mirkin et al. 1996, in which 13nm diameter gold spheres were attached to 8-base DNA oligomers and then the resulting "hairy spheres" colloid could be made to agglomerate into a solid mass by the addition of unbound longer DNA oligomers, with ends complementary to the 8-base bound sequences, to the solution. However, no effort was made to precisely control this agglomeration either at the micro scale (i.e. how many DNAs were attached to each gold sphere) or at the macro scale (i.e. control which spheres we want to attach to which by choosing what DNA sequences to attach to them). In the same issue of *Nature* is an article by Alivisatos 1996 and a "News and Views" column (page 581) on the same sort of subject. The idea is definitely there that somehow this kind of thing is leading toward nanotechnology, but no precise algorithmic strategy is elucidated.

Finally, the paper which almost reaches our idea is Shi & Bergstom 1997. In this paper, it is proposed that 4 DNA arms could be attached to a rigid tetrahedral hub, and the resulting objects could be joined to make nanostructures – same idea as mine. And then Shi & Bergstrom actually carried out a simplified version of their plan, in which they used hubs with *two* arms at a 120-degree angle, and only using *two* DNA sequences. There are several ways in which Shi & Bergstom fell short of my proposal, however. First, although their arms were at 120-degree angles, their monomers were found to self-assemble into 2-gons most commonly, also with 3,4,5,6, and 7-gons being produced with decreasing frequency. I would regard this as an experimental failure from the standpoint of reliable link-forming for building large structures – because the 4 most common products should have been impossible. Second (and this is related), their arms were single stranded DNA, and not double stranded DNA with sticky ends. Thus, ligation, and therefore permanent covalently bonded attachment, was for them impossible. Third and finally, an explicit assembly

[20] Some other ideas were needed to accomplish topological control over knots by the use of "Z-DNA" as well as regular DNA, and to activate various reactant strands with the aid of restriction enzymes, but these ideas don't seem relevant for us.

[21] Topological linkage of all strands will also occur, but since all our DNA strands will also be linked by covalent bonds, this is not of great interest.

algorithm assuring logarithmic synthetic depth (like ours based on the centroid decomposition of spanning trees) was not proposed.

It seems to me that the only real defect of the Shi & Bergstrom work, the rigidity failure, might be susceptible to an easy fix.

Mentally approximate double stranded DNA as a "rigid rod" (cf. §8.1). If you attach a rigid rod via just 1 single bond to something, which is what Shi & Bergstom did, it is free to rotate; imagine the rigid rod sweeping out a cone as it rotates attached to the ground at some angle. But, if you attach it to the ground via *two* free-to-rotate bonds, each at any location and having any rotational axes whatever, then the only degree of freedom is removed and the result can't move. So in principle, our proposal involving *double stranded* DNA arms attached to their hubs by *two* bonds, combined with a final ligation step so that all "loose ends" are eliminated, will solve this problem.

7.3 Other possible primitive operations

Additional optional primitives could result in more interesting capabilities.

For example, suppose some of the arms of some of our building blocks incorporated "fusible links." Then by application of the appropriate reagent we could perform

Operation C [cut]: Cut all fusible links.

This would allow the synthesis of, for example, a statue of liberty with rotating gears and cogs and sliding rods inside.

Another possibility is as follows. At some future time, it may be that a large number of useful macromolecular "parts" (e.g. gears, rods, electrical switches,...) may become available, because of the techniques above. We would then like to glue the parts together to construct a nanomachine. This will require standard chemical "interfaces" between such parts so that they may be joined again using appropriate reagents. The construction process requires operations just like the ones we've been discussing, except that we will now be at a "higher level" of the "size heirarchy," and each node in our trees is really going to be a macromolecule itself, which had been constructed by operations at "lower levels." At high levels, it may be that the joining operations and so forth could be of a completely different character (i.e. depending on non-covalent bonding, e.g. see Anelli 1992, Annabilino 1995) than at lower levels. Anyhow, if things at low levels need to be designed to permit future additional operations to happen at high levels, that will make the chemist's job even more difficult.

7.4 "Directed tree" viewpoint removes the need for stacked protection

Although the repertoires we've suggested above seem to be fairly small, and it is easy to find chemical reactions which implement all the functionalities we want... it is still very difficult to design molecules which support *all* of our desired functionalities, compatibly, while still staying rigid and stable, at the same time!

So here is an additional simplifying idea designed to make that task more possible: Instead of trees, think *directed* trees. Namely, every edge in a tree may be given a direction arrow pointing toward the root (and any node may be designated "root"). This view will enable us to eliminate the need for stacked protection (operation P_{kn} with $n > 1$) and only require 1-level protection of the rootward-pointing reactive groups on the rootward reactant.

The price we have to pay for getting rid of stacked protection is that we lose the guarantee that our entire synthesis tree will only have depth $O(\log N)$ where N is the number of nodes in the (tree shaped) goal molecule. Instead, in the worst case $O(D)$ steps will be required, where D is the graph theoretic diameter of the goal molecule (again we are assuming that assembleability problems as in §6 never have any significant impact). However, if the goal molecule is fairly "round" and if N is not huge, this may not make a great deal of difference because D will be of order $N^{1/3}$, which does not grow much faster than $\log N$ when $N < 10^4$.

The idea is this. At a node η in the tree, there are generally ≤ 2 subtrees (in the leafward direction) and 1 parent node (in the rootward direction). We are going to do a "rooted tree assembly" as defined in §6.2.

So consider a "tree node" building block with three hands T, A, and B, where T is intended to point toward the root and A and B toward the leaves of the tree we are building.

```
      T
      |      Tinkertoy for use
      *       as node in directed tree.
     / \       T points rootwards.
    A   B
```

Operation set #0: Join A, join B, activate T. ("Join i" will take a tree with exactly one activated hand, of type T, and join it to a new node [whose T arm is not activated] at a hand of type $i \in \{A, B\}$.)

Theorem: Operation set #0 and an unlimited supply of tree nodes as above (as well as tree nodes with amputated limbs that can serve as "end caps") will suffice to build any rooted tree graph structure whose nodes have degree ≤ 3. (Also by making k kinds of "legs," $k \geq 2$, we can make arbitrary trees with branching factor $\leq k$.) □

Here are some variants of this 4-operation set, which work just as well:

Operation set #1: protect A, deprotect A, protect T, deprotect T, join. ("Join" will take a tree with exactly one unprotected hand, of type T, and join it to a new node with exactly one unprotected hand, of type A or B. Actually only the deprotection operations are needed, if the tinkertoys are assumed initially provided with T and A pre-protected.)

Operation set #2: Root-activate^A, root-activate^B, join. (Root-activate^i will convert any hand of type T to a form which is capable of being joined to a hand of type $i \in \{A, B\}$. Join will take a tree with exactly one free hand, of type T^i, and join it to a new node at a hand of type i.)

Operation set #3: Activate A, join A, join B, protect T, deprotect T. ("Join B" will take a tree with exactly one free hand, of type T, and join it to a new node at a hand of type B. "Activate A" will take a tree with no hands of type B, and convert any hands of type A to an active form. "Join A" will take a tree with exactly one free hand, of type T, and join it to a hand of type active-A; there are assumed to be no hands of type B. Actually only "deprotect T" and not "protect T" is needed, if the tinkertoys are assumed initially provided with T pre-protected.)

Operation set #4: Activate A, join A, join B, activate T. ("Join B" will take a tree with exactly one free hand, of type T, and join it to a new node at a hand of type B. "Activate A" will take a tree with no hands of type B, and convert any hands of type A to an active form. "Join A" will take a tree with exactly one free hand, of type T, and join it to a hand of type active-A; there are assumed to be no hands of type B.)

8 Molecular rigidity and stability

8.1 Rigidity and its limits

We want to synthesize rigid molecules so that we can completely control geometry. How can we understand rigidity?

The "ball and stick" mental model, involving fixed bond lengths and angles, is a good guide about molecular rigidity but is not always a sophisticated enough model to give you the answer.

In cases where the ball and stick model is suspect, one must fall back on experimental data about interconversion rates and/or theoretical calculations of rotation barrier energies. The true answer is never absolute rigidity or complete flexibility as in the ball and stick picture. Instead, molecules are more or less rigid depending on the temperature, the barrier energies, atomic bond spring constants, and upon the energy differences among various alternative forms.

Also, it isn't clear how much rigidity we really require. That depends on the purposes we intend to use our nano-constructs for. For some purposes it may suffice that our molecule be in the conformation we want some high percentage of the time, e.g. 99%, even though it frequently makes excursions to temporarily explore other, less favored, conformations and also oscillates wildly within some range within each conformation. For other purposes, we might want to demand that it *stays* in *one* conformation for a long time, e.g. a year. Most strongly of all, we might want to demand further that all the amplitudes of all vibrations be below some bound. Because conformation lifetimes and vibration amplitudes depend on temperature, extra rigidity is obtainable (post-synthetically) by cooling. Indeed at 10 Kelvin, all of the bonds we discuss below will be highly rigid; even rotation about C-C bonds is eliminated.

Available thermal energies: at T =room temperature, $k_B T \approx 0.59$Kcal/mole.

A well known example of failure of the ball and stick picture is: "inversion" of NR_3 "pyramids" occurs rapidly at room temperature due to the low barrier $\approx$ 6Kcal/mole.

On the other hand, the corresponding barriers for phosphines (29Kcal/mole), arsines (42Kcal/mole), and for chirality inversion about 4-valent C's or Si's (80Kcal/mole?) are much larger and these effects almost never occur.

For another example, the two chair forms of cyclohexane theoretically should be completely rigid if their bond lengths and angles were fixed[22]. But in reality, bonds can stretch and angles can bend slightly, so that in fact cyclohexane interconverts to two "twist-boat" forms (each 5.5 Kcal/mole less stable than the chair) by surmounting a 10.8Kcal/mole energy barrier. At room temperature, such barrier crossings happen 10^4-10^5 times per second. Thus for our purposes cyclohexane seems insufficiently rigid. On the other hand, only $\approx 10^{-4}$ of the cyclohexanes are present in the twist-boat form at any time. That might argue that cyclohexanes *are* sufficiently rigid. Structures with large numbers of cyclohexane chairs (such as diamond) are rigid; but they are hard to synthesize.

Rotational barriers for the C-C single bond are 2.9Kcal/mole for CH_3-CH_3, and higher barriers if the substituents are larger than H's, for example CH_3CH_2-CH_2CH_3 has a barrier height of 3.4Kcal/mole, CH_3-$C(CH_3)_3$ has 4.7Kcal/mole, and *ortho* $(CH_3)_3$C-PhBr has 9.1Kcal/mole. In all these cases, at room temperature we expect

[22] WLOG the first 3 carbons are fixed at (0,0,0), (1,0,0), $(1-c, s, 0)$ where $c = \cos(\text{bond angle})$, $s = \sin$. The remaining 3 carbons have $3 \times 3 = 9$ degrees of positional freedom, but they are constrained by 4 length equations (4 bonds, length= 1) and 5 bond angle equations, i.e. 9 constraints total. Since $9 = 9$ we have total rigidity.

essentially free rotation, i.e. once in every 100-4 $\times 10^6$ barrier crossing attempts, which at a 20 GHz attempt frequency ranges from 200 MHz down to 5 KHz. (The Arrhenius probability that a barrier of height E will be crossed on an attempt is $\exp(\frac{-E}{k_B T})$.)

This contrasts with the C=C double bond, which won't rotate at common temperatures due to the huge rotation barrier of 62-65 kcal/mole. (At an attempt frequency of 10GHz, the barrier will be surmounted only after $\approx 10^{28}$ years!)

However, C=N and N=N bonds have much smaller barriers to rotation. C=N bonds typically have rotation barriers ranging from 15-40 kcal/mole, and N=N bonds 23-36 kcal/mole, and the high ends of these ranges are only achieved when strongly electronegative groups (such as O, F) are attached to the nitrogen atoms; otherwise we stay in the lower 5 kcal/mole of the given ranges. The corresponding room temperature halflives range from 10 sec to 10^{12} years.

Even for C=C double bonds, however (in fact for all bonds), elastic deformation is possible. Typical force constants are: for C-C stretching: 5 N/cm; for C=C stretching: 10 N/cm. To bend a C-C bond a small angle θ requires energy $\kappa\theta^2$, $\kappa \approx .6$aJ/radian^2. The torsional energy constant for a C=C bond may be estimated from the known barrier energy (torsional energy at 90°). From these numbers, we conclude that at room temperature, C-C bonds are typically (assuming the root mean square stretching energy is $k_B T/2$) stretched or contracted by 2% and bent by 5°. Typical torsion angles for C=C bonds at room temperature are 10°. At temperatures x times smaller, these distortions will be $\sqrt{x}$ times smaller.

Thus, even "rigid" molecules are fairly floppy at room temperature. So it would seem that for many purposes one needs more rigidity than can be obtained from any tree structure, no matter what sort of bonds it is made of. You need a lattice-like structure instead, or at least the "bonds" in the tree structure would really need to be bundles of bonds. However, we are essentially forced to use trees during the intermediate stages of our synthesis, otherwise the chemistry would be impossibly difficult! Also bond bundles are rather difficult to make en masse! (We do try to make amends at the end of our synthesis by using "rigidizing reagents," and possibly these could be used at earlier synthetic stages in some cases, but this may be too late.) If we are determined to use plain trees (which we are) then it is certainly best if one's trees have as small graph theoretic diameter as possible – which is a measure we *are* taking.

8.1.1 Atropoisomers

More interesting cases are C-C (or other) single bonds with slight double bond character due to resonance stabilization effects, or having rotation hindered by the presence of large nearby groups. Will they rotate? In the comparatively rare cases (discussed below) when rotation is prevented, we get "atropoisomers."

Normally, resonance stabilization is not enough to prevent C-C bond rotation. For example, the C-C single bond in H_2C=CH-CH=CH_2 is about 2-3 Kcal/mole stronger than usual due to resonance stabilization from the conjugated double bonds[23]. It also has a rotational energy barrier of about 3.9Kcal/mole, which is also larger than a typical C-C bond, but this still is not enough to prevent rapid rotational transitions at room temperature. It is, however, enough to make butadiene, and similar molecules, appear planar and "trans" in X-ray structure determinations since the barrier to rotation is surmounted well under 1% of the time at room temperature and since the "cis" form is sterically disfavored. This is especially true in cases where the double bond is conjugate to an aromatic ring. For example, none of the double bonds in hexa vinyl benzene Ph$(CH=CH_2)_6$ are attacked by bromine, which is presumed to be because they all lie close to the plane of the ring almost all of the time because the Ph-Vinyl single bonds rarely rotate, and hence steric hindrances prevent the bromines from maneuvering into attacking positions.

The exception is the "peptide bond" in amides (and thioamides) which is the type of single bond, among those commonly encountered, with the greatest double bond and resonance stabilized character. As evidence for this, the nitrogen in amides is planar instead of pyramidal, and the bond length is 1.32 instead of the usual 1.47Å for a C-N bond. It is commonly said that the peptide bond has "40% double bond character."

Thus these two thioamide compounds

[23]This compares with 4-6 Kcal/mole of stabilization energy per bond in *aromatic* ring compounds. Really, though, bonds in aromatic rings are much more stable and resistant to distortions that this number might imply, because destroying just one bond breaks the aromaticity of all the bonds in the whole aromatic ring.

```
   Ph                     Ph
     \                      \
      C=S                    C=S
     /      (a)             /       (b)
    N                      N
   / \                    / \
  CH  CH - Ph      Ph - CH  CH
    3   2                 2   3

        Two atropoisomers
```

only interconvert with a halflife of days in solution at room temperature. A review is (Oki 1983).

The peptide bond in $CH_3CO\text{-}N(CH_3)_2$ can rotate, but there is a barrier of 13.95Kcal/mole which tends to keep it planar (Johnson 1997). Consequently, interconversion between *cis* and *trans* forms occurs with a typical lifetime of about 1 second at room temperature, with the "attempt frequency" being about 16 GHz. In other words, it is planar except for about 1 time in 10^{10}, when it gets to flip over.

Other amides (tables are in Isaacs 1995 and Eliel & Wilen 1994) exhibit larger rotational barriers, ranging from 18-25 Kcal/mole, implying room temperature lifetimes ranging from 10 minutes to 5 years; and the barriers in thioamides typically seem about 7 Kcal/mole larger than amides, implying far longer lives.

Rotation of C-C single bonds can also be sterically hindered by bulky groups. Another example from the Oki review is

```
     H C
      3 \  NO
         O \ 2
        /   \
    HC-C     C-CH
   //  \\  //  \\    (note, the righthand benzene ring
  HC     C-C     C    is rotated perpendicular to the plane
    \   /   \   /     of the paper, in which the lefthand
    HC=CH    C=CH     benzene ring lies.)
            /
        HO-C
            \\
             O
```

which racemizes with a halflife of 9.4 min in ethanol at room temperature. With 4 large ortho groups instead of just 3, (see March page 101), rotation is essentially prevented. Triptycene derivatives can also exhibit similarly hindered rotation; see March, compound #102 on page 162, to find examples with barrier heights ranging from 10.3 to 17.6Kcal/mole. (Even more amusing are "geared" pairs of bonds which must rotate in concert due to bulky side groups acting like the teeth of gears.)

"Hydrogen bonds" can also prevent or hinder rotation about single bonds because such rotation would have to break the hydrogen bonds. The energies of hydrogen bonds are 3-6Kcal/mole.

8.1.2 Conclusion

So, in conclusion, to keep our molecules rigid it seems best to rely on aromatic rings and double bonds, which cannot rotate. Naturally more bonds (i.e. by use of multi-ring structures instead of trees) will imply more rigidity, but also incur more synthetic difficulty. In the cases when we do use apparently rotatable single bonds, hopefully rotation can be sufficiently reduced with the aid of steric hindrances, hydrogen bonding, or by using peptide bonds and bonds with partial double-bond character.

One way to quantify rigidity is to consider the worst case, which is a "fiber" made by linking n tinkertoys in a row, and asking what its "persistence length" will be at room temperature.

Definition: The "persistence length" of a polymer chain is the limiting expectation value of the projection of the endpoint of an n-long chain, on the direction of the first segment (which starts at the origin) as $n \to \infty$. This is also called the "Kuhn length."

Example: a random walk in a cubical grid has Kuhn length equal to the square-side length. Presumably tinkertoy construction plans won't suffer from non-rigidity provided we stay well below this persistence length.

For typical flexible chain polymers (e.g. polyethylene $(CH_2)_n$) the Kuhn length ℓ is 10-30$\mathring{A}$ at room temperature, e.g. about 7 monomer lengths. This is useless for tinkertoys.

DNA has $\ell = 500\AA$, an exceptionally large figure for a "flexible" structure (Hagerman 1988; this figure depends heavily on ion concentrations and temperature). This is due to the presence of two intertwined helices linked by hydrogen bonds, a large helix diameter ($20\AA$), and the "stacked plates" arrangement of the bases inside the helix. Since each base pair in DNA requires about $3.4\AA$ of axial length, this is equivalent to about 147 monomer units, or to about 14.7 helix turns, since there are 10 base pairs per helix turn.

For polymers consisting entirely of bonds with noticeable partial double bond character (e.g. amide bonds and bonds conjugated between amides and/or aromatic rings), the Kuhn lengths are considerably longer. Thus "polycarbamate" $(\text{CO-NR})_n$ (this has also been called "poly(R-isocyanate)" and "nylon 1;" here R is a hydrogen or short alkane) has $\ell = 1000\AA$ (500 monomer lengths) when R is H (Tsvetkov & Shtennikova 1978), and $\ell = 410 \pm 15\AA$ when R is C_6H_{13} (Rubingh & Yu 1976). "Kevlar" and "PBA"

```
           H
           C
          / \\    H     H
        HC    C--N      C
(    H   ||   |   \   / \\        )
      N--C    CH   C--C    CH      n
     /    \ //     || ||   |
 ---C      C       O  C    C---
   ||      H           \ //
    O                   C
                        H
```

Kevlar.

```
        H
        C
       / \\    H
     HC    C--N
(     ||   |   \      )
   ---C    CH   C---   n
       \ //     ||
        C       O
        H
```

PBA (Poly benzamide)

have $\ell = 1300 \pm 130\AA$ (83 monomer lengths for the monomer shown) and $\ell = 2100\AA$ (320 monomer lengths) respectively (Tsvetkov & Shtennikova 1978). This is probably good enough for many tinkertoy purposes.

However, the "rigid rod" character of these polymers depends upon the fact that amide bonds and bonds conjugated between amides and/or aromatic rings are harder to rotate than a normal C-C bond. Presumably, for fibers made of rigid units linked by *truly* rigid bonds (e.g. fused aromatic rings as in §13) of the sort we are preferring, or solely by bonds whose rotation is irrelevant for rigidity, one should expect much greater rigidity, perhaps 4-100× larger ℓ, i.e. .5-20μm. In fact, several such "true rigid rod polymers" have been made (Yang 1989)[24] including polyimides, the "BBL ladder polymer," and Poly(*p*-phenylene-*trans*-benzobisthiazole) (PBT, also sold commercially as "Aftech II"). PBT fibers have the largest known elastic modulus among all organic polymers.

```
            H            H
        N   C            C
      // \ / \\         / \\
( ---C    C    C--S   HC    C--- )
      \   ||   |   \   ||   |     n
       S--C    C    C--C    CH
          \ // \ //    \ //
            C    N       C
            H            H
```

Poly(p-phenylene-trans-benzobisthiazole), or PBT: a "rigid rod" polymer

Unfortunately, Crosby et al. 1981 found the persistence length of PBT in chlorosulfonic acid at room temperature is only $\ell = 640 \pm 90\AA$ (which agrees reasonably well with a recent theoretical estimate of 515-554$\AA$; [Mattice & Zhang 1993] in thermal vacuum) and ascribe to another author the value $\ell = 1500\AA$ to the BBL ladder polymer in methanesulfonic acid – and both of these values lie below the $\ell = 2100\AA$ value for the supposedly less rigid PBA![25]

[24] Going beyond these are carbon "nanotubes" (Ebbesen 1996). Nanotubes are really 2-dimensional structures ("cylindrical graphite") and hence are not very relevant to our present comparison of one dimensional fibers. (Naturally tinkertoy structures may be made more rigid by constructing anything other than a 1-dimensional fiber.) Tinkertoys *made* of nanotubes might be great, except that nobody knows how to controllably make nanotubes. Anyhow, a carbon nanotube 5000nm long, and with inner radius 1nm and outer radius 8nm, was observed by Treacy et al 1996 to be oscillating thermally with an RMS tip amplitude of 10nm at room temperature (and with mean square amplitude linearly proportional to temperature). This leads me to estimate that at room temperature, if the nanotube they observed had somehow been extended to be very long, its persistence length would have been about 3 meters! (Derivation: assume an RMS curvature of .002 radians per 5000nm length and all such curvatures on disjoint 5000nm segments are independent. To reach a curvature of $\pi/2 \approx 1.57$ radians we'd need to adjoin typically $(1.57/.002)^2 \approx 6 \times 10^5$ such segments, whose total length is then 3meters. Even in the extremely conservative model in which all curvatures on 5000nm segments are assumed to be in the same direction, we still predict a macroscopic persistence length of $> 5000 \times 785$nm= 3.9mm. Of course, these persistence lengths are far longer than any nanotube that has ever been made.

[25] Crosby also ascribes to other authors in 1976 the measurements $\ell = 240\AA$ *and* $\ell = 400\AA$ for PBA, which contradict Tsvetkov &

On the other hand, it is possible that the severely acidic solvents used by these experimenters de-aromaticized the rings by protonation, reducing rigidity. An X-ray diffraction study on one of these materials ("PBO") in the solid state (Adams et al 1989) suggests a persistence length under non-protonating conditions that indeed is $> 5000Å$.

8.2 Stability

We want our nano-molecular creations to be stable molecules. The "stability" of a molecule depends upon the conditions it is exposed to and just how much and what kinds of degradation are acceptable. It may be quantitatively expressed as a halflife.

I know of no sure way to predict the halflife of a substance from its structural formula; the only way to determine it is experiment. Extensive experiments have been conducted on dyes and pigments to determine their stability to sunlight, acids, alkali, organic solvents, bleach, and laundering; leafing through the "Colour index" to see which formulas are more stable is an educational experience. Chemicals which have been around for hundreds, or even better millions, of years are an even better bet.

But certainly there are some rules of thumb one should obey if one is in the business of designing stable molecules. First, stay with high stability bond types such as C-O, C-H, C-C, C-F. Second, avoid known reactive and vulnerable functional groups such as aldehydes -CHO, hydroxyl amines, nitro groups, alkyl halides C-I, C-Br, C-Cl, and non-aromatic double bonds C=C (which can be attacked by atmospheric ozone and which have other nasty habits). (Very reactive groups such as acyl chlorides, azides, peroxides, etc. are of course in a different ballpark entirely and should never be considered.) Third, having high bond dissocation energies does not necessarily mean anything if it is possible to perturb the molecule to get even higher ones. Thus, obviously, avoid bond strain such as in 3 or 4 membered rings.

Conversely, having low energies also may not mean anything if such perturbations are very difficult. Aromatic molecules have a lot of stability because destroying a bond requires breaking the aromaticity of the entire ring. I recommend them. The most inert ways to incorporate oxygen and nitrogen are undoubtably ethers, amides, and inside aromatic heterocycles such as pyrrole.

Clar 1972 has made an extensive study of the stability of polycyclic fully aromatic systems of hexagons. One can draw a "Clar circle" inside some of the hexagons to represent aromatic sextets. Hexagons that share an edge are not allowed to both have Clar circles. The more Clar circles that may be drawn, the more stable the compound, and Clar concludes that the most stable such compounds are the "all benzenoid systems," in which one may draw Clar circles in such a way that every carbon is a member of some Clar circle. For example, triphenylene (§9.1) and benzene are ABS's, but napthalene and anthracene (the polyacenes, see below, with $N = 2$ and $N = 3$) are not. ABS's tend to be pale yellow or colorless, do not dissolve in concentrated H_2SO_4, do not react with maleic anhydride, and have extremely great thermal stability (e.g. the 6-fold symmetric "hexa(peri)benzocoronene" $C_{42}H_{18}$ remains a stable solid beyond the temperature at which glass melts).

Clar's theory can have dramatic consequences for the putative design of rigid, fully aromatic nanocomponents, as was pointed out to me by T.W.Ebbesen. Thus the N-ring polyacenes become increasingly unstable and reactive for larger N.

```
 / \ /\\ /\\ /\\ /\\ /\\ /\\ /\\ /\\ /\\        "Polyacene." Low aromatic
| O |   |   |   |   |   |   |   |   |   | ...    stabilization since only
 \ / \// \// \// \// \// \// \// \// \//      one Clar circle may be drawn.
```

When $N = 7$ we have heptacene (Clar 1964, 1972), which has been synthesized but is too unstable to prepare in the pure state, and according to Clar "considering the extreme reactivity of heptacene, the possibility that the synthesis of octacene could succeed seems remote." Indeed, although polyacene derivatives (alcohols, quinones) have been synthesized up to $N = 11$, all attempts so far to reduce these to pure polyacenes have failed.

But the situation seems entirely different with the following series.

```
     / \         / \         / \
    | O |       | O |       | O |
 / \ / \ / \ / \ / \ / \ / \ / \ / \ / \        "Bumpy polyacenes."
| O |   |   | O |   |   | O |   |   | O | ...      Each are Clar
 \ / \ / \ / \ / \ / \ / \ / \ / \ / \ /       "all-benzenoid systems."
    | O |       | O |       | O |
     \ /         \ /         \ /
```

The first few bumpy polyacenes (triphenylene, dibenzopyrene, tetrabenzanthracene, tetrabenzopentacene, and tetrabenzoheptacene [$C_{42}H_{22}$, melting point 485C]) have been prepared and all seem extremely inert and stable.

Shtennikova 1978 as well as each other!

Unfortunately Clar and his successors (e.g. see Gutman & Cyvin 1989) did not extend their observations to *heterocyclic* fully aromatic ring systems. But in the absence of evidence to the contrary, presumably the most stable compounds are "ASSes" (All Sextet Systems!) in which every carbon or hetero atom is a member of a ring with a wholy owned sextet of π electrons.

I have followed all of the guidelines above in my molecular designs presented below, with these two possible exceptions:

1. I have used azo and conjugated double bond linkages between aromatic rings (Ar-N=N-Ar and Ar-CH=CH-Ar) in[26] §10. Although the azo linkage consists of comparatively low energy bonds, there is an aromatic stability enhancement. This is presumably why the azo linkage has been used extensively in dyes and pigments, some of which are known to last for hundreds of years.

2. The fully aromatic tinkertoy constructs of §13 are all benzenoid Clar-optimal structures, *except* that the "indolo-indole" linkages each incorporate one "fixed" C=C double bond that is not part of the maximal Clar sextet system. These could be vulnerable[27].

9 Some useful molecules and reactions

9.1 A few molecules with interesting geometrical properties

Here is a brief discussion of 4 rigid symmetrical molecules that may be useful for building tinkertoys around. All of the below molecules are rigid, and variants of the first three have been synthesized in which the outer benzene rings are replaced with outward pointing napthalenes.

```
           H
           C
         // \
       HC    CH
   H    |    ||
   C    C    CH
  / \\ / \\ /
HC    C    C
 ||   |    |
HC    C    C
  \ // \ // \
   C    C    CH
   H    |    ||
       HC    CH
         \\ /
           C
           H
```

Triphenylene

```
    H            H
    C            C
  // \         / \\
HC    CH     HC    CH
 |    ||      ||    |
HC    C    N   C    CH
  \\ / \ // \ / \\ /
    C   C    C    C
    |   |    ||   /
    C---N    N--C
  //     \  .    \\
  N       Cu      N
   \      .  \    /
    C===N    N--C
    |   |    |   \\
    C   C    C    C
  // \ / \\ / \\ / \
HC    C    N    C   CH
 |    ||        |   ||
HC    CH       HC   CH
  \\ /           \\ /
    C              C
    H              H
```

Copper Phthalocyanine

Triphenylene might be a good basis for a planar tinkertoy with 3-fold symmetry. A large number of derivatives are known; a survey is Buess & Lawson 1960. The long term stability of triphenylene is proven by the fact that it occurs naturally in coal tars; also it is thought to be the most stable isomer of $C_{18}H_{12}$.

Copper Phthalocyanine has perfect square symmetry and is wholy planar. It is produced in 10,000-ton quantity every year because it happens to be a powerful blue pigment. It also is exceptionally stable (i.e. stable to alkali, acids, sublimes unchanged at 580°C in vacuum; generally regarded as the most stable organic colorant) and cheap. Over 5000 papers have been written on it and thousands of derivatives are known. It also has interesting semiconductive and photoconductive properties due to the fact that it is wholy aromatic. A survey is Thomas 1990. Although this

[26] An alternative, and likely better, method was suggested there which should enable avoiding them.

[27] However, as was mentioned in §13, it is probably possible to redesign the "indolo-indole" linkages in favor of "carbazole" linkages, which would presumably be more stable.

molecule seems perfectly suited as the basis for tinkertoys for the production of *planar* structures, it is conceivably possible to extend into the third dimension too by "stacking" phthalocyanine "plates." One replaces the Cu by an Si and links the Si's in chains such as Si-O-Si-O-Si... vertically. (Also, Si-O-C_2H_4-O-Si-O-C_2H_4-O-Si and so on have been used. One prepares HO-R-O-Si-O-R-OH or HO-Si-OH phthalocyanine and then dehydrates it.) However, it is not clear to me that the plates in such stacks can be prevented from rotating.

```
Triptycene      1,3,5,7 adamantane tetra carboxylic acid
```

Triptycene (surveyed in Skvarenko et al. 1974; a synthesis as an undergraduate experiment is in Feiser & Williamson 1992) has 120-120-120 3-fold symmetry. Many derivatives are available, and attachments are possible at the ends as well as the sides of each of the 3 phenyl "arms" as well as above and below the axial carbons and thus into the third dimension. Although the axial connections are normally free to rotate (e.g. if you just attach via a single bond to carbons 9 and/or 10) some compounds have been made with a 5-member ring which includes carbons 9 and 8 (and the un-numbered one between) from triptycene, plus two additional carbons. Such structures cannot rotate and are fairly rigid although perhaps not perfectly so.

A convenient synthesis of 1,3,5,7 adamantane tetra carboxylic acid was mentioned by Newkome et al. 1992. Adamantane (same molecule with the COOH moieties removed) has been found in some petroleums, which proves its stability. A review of adamantane chemistry is Bingham & Schleyer 1971. Although as we mentioned chair cyclohexane is not rigid at room temperature, adamantane is much more rigid since it incorporates three chair cyclohexane rings. It has the symmetry of a regular tetrahedron.

9.2 Brief review of some cyclization reactions

Because of our desire for structural rigidity and the fact that top quality rigidity requires rings (§8.1), it seems the best kind of joining reaction would be a "cycloaddition" or "cyclocondensation." The three such reactions that look most interesting for our purposes (because one can exert the most control over them) are the Fischer Indole Synthesis,

```
    H                                                   H
    C               R                                   C       R
  // \          H\ /    deammonatory &                // \      /
HC    CH        H-C     dehydratory conditions      HC    C---C
 |   ||      +    |    ------------------------->    |    ||   ||
HC    C   NH    O=C         ZnCl  catalyst          HC    C   C
  \\ / \ /  2      \             2                     \\ / \ / \
    C   N           R'                                  C   N   R'
    |   H                                               H   H
    CH
      3

      Fischer Indole Synthesis.
The purpose of the methyl group on the phenyl hydrazine is
to force the indole ring to attach at the carbon above the
hydrazine group. If R' is a carbon with less than 2 H's on it, or if R' is H,
then we similarly force specificity for the carbon above the ketone group.
```

the Robinson Annulation,

```
    *     HCH  R                                   *   *   R
   / \     \\ /     Basic & dehydratory           / \ / \ /
  *   *      C      conditions, catalyst        *   CH  CH
  |   |   +  |    -------------------------->   |   |   |
  *   C=O    C=O                                *   C   C=O
   \ /      /                                    \ / \\/
    C    H C                                      C   C
   / \    2 \                                    / \   \
H C   CH     R'                               H C   CH  R'
 3      3                                      3      3
           A vinyl
A cyclic   ketone.       ("*" denotes a CH  unit.)
Ketone.                                  2
                 Robinson Annulation.
```

The purpose of the 2 methyl groups on the cyclohexanone is
to force the ring to attach at the carbon above the ketone.
By use of a chiral catalyst mixture involving L-proline
it is possible to exert some stereochemical control, but
this of course is irrelevant if one plans to aromaticize the rings.

and the Diels Alder cycloadditions.

```
         R   R'                                R   R'
          \ /    ("*" denotes a CH  unit.)      \ /
    *      C                    2             *   C
   / \     \\                              / \ / \
  *   CH     CH                           *   CH  CH
  |   ||  +  |     ------------------->   |   |   ||
  *   CH     CH      MgBr  catalyst       *   CH  CH
   \ /     //           2                  \ / \ /
    C   R"-C                                C   C
   ||      |                               ||   |\
    O      S-Ph                             O   S R"
                                               /
        A Diels Alder cycloaddition.          Ph
```

The purpose of the phenylthio group is: it exercises a strong
directing effect to prevent the ring from forming with the right hand
side turned upside down. This group can easily be removed later.
There are several alternative directing methods, see Carruthers 1990.
There are also ways to exert stereochemical control, but these
of course are irrelevant if one plans to aromaticize the rings.

Of these, the Fischer indole synthesis seems to be the most promising because it fuses a fully aromatic (pyrrole) ring onto another fully aromatic (benzene) ring forming a completely rigid and planar structure in a one-pot reaction. Unfortunately, the yields seem[28] to be rather low, 50-70%.

The other two reactions can also be used to create aromatic rings, but only by appending extra reaction steps designed to aromaticize the rings. Normally they create cyclohexane/ene rings, which as we've mentioned are comparatively nonrigid at room temperature. A review of the F.I.S. is Robinson 1988.

10 A tinkertoy design based on stacked protection and deprotection schemes

How can we do stacked protection, as in operation P_{kn}, with deprotection as in operation D?

I have searched the chemical literature in vain for "stacked" protection schemes, so apparently this is a new idea. We will devise one plan for such a scheme by combining polypeptide and activated aryl chemistry.

[28]The yields for the other cycloadditions aren't so great either. Yields mentioned in the Gawley 1976 and Jung 1976 reviews of the Robinson Annulation are typically 34-85%, although in several cases yields are > 90%. Diels Alder cycloadditions are reviewed by Carruthers 1990 who mentions yields ranging 43-80%.

10.1 A preliminary: Some peptide chemistry

There are several well developed methods for the synthesis of polypeptides, or their extension at either end. We will now describe a simple version of one such method, because it makes a nice introductory example of our tree-like synthesis strategy, particularly simple because the things we are synthesizing (polypeptide chains) are only one dimensional.

The plan to produce an N-long polypeptide chain is to recursively synthesize the left half and then right half and then join them. Call the left and right halves "A" and "B." Since polypeptide chains have amine and carboxylic acid moieties sticking out of their ends, we need a reaction sequence that accomplishes either this

```
H N-A-COOH  +  H N-B-COOH  --->  H N-A-CO-NH-B-COOH
 2              2                 2
```

or this

```
H N-B-COOH  +  H N-A-COOH  --->  H N-B-CO-NH-A-COOH
 2              2                 2
```

which is of course the same thing but with the names "A" and "B" mentally exchanged, so we need only to describe the former transformation.

We first *protect*[29] the amine group on A for example by reaction with benyl chloroformate Ph-CH_2-O-CO-Cl to yield the benzyloxycarbonyl amide Ph-CH_2-O-CO-NH-A-COOH.

Next, we *activate* A's carboxylic acid by either (1) reacting it with $SOCl_2$ to convert it to an acyl chloride B-COCl (this activation method is not often used in polypeptide synthesis for biological purposes because it can catalyze racemization of the activated amino acid, but it is the most common activation method elsewhere in organic chemistry and if chirality is unimportant to us, as it is here, this is fine), or (2) by use of the special reagent "DCC" (dicyclohexylcarbodiimide), with certain racemization-inhibiting additives, or (3) reaction with pentachlorophenol $HOPhCl_5$ to create a highly reactive ester, or (4) there is a fourth alternative based on azide chemistry.

Finally, we react the activated and protected A with B to form the twice as long peptide chain AB. (The DCC activation method has the advantage that it may be done simultaneously with the AB joining reaction in one pot, but in that case, B's carboxyl moiety would also have to be pre-protected, for example by conversion into its benzyl or tert-butoxy ester. This is to prevent B's DCC-activation allowing the formation of unwanted polypeptides BB.)

The benzyloxycarbonyl amide protection of the amine on A's left end may now be removed by either Pd-catalyzed hydrogenation or hydrolysis by HBr in cold acetic acid. (The purpose of the initial protection of A's amine moiety was of course to prevent the self reaction to form AA. Certain amino acids have reactive side groups which also have to be pre-protected before they can be used in this synthetic process, see Mackie et al. page 357.)

The entire synthesis of an N-long polypeptide chain by this algorithm requires $\log_2 N$ of these joining stages, and assuming 90% yields at each joining stage, the overall yield will be $N^{-.152}$ (here $\log(.9)/\log(2) = -.1520...$).

For the *removal* of one amino acid at a time from the amine end of a polypeptide, we have the *Edman degradation*: apply Ph-N=C=S (phenyl isothiocyanate) in an alkaline medium, followed by hydrolysis with trifluoroacetic acid. (More details about the many chemical and mechanical refinements of the Edman degradation that have brought its typical yield per step up to 96% for unskilled [and 99% for especially careful] operators may be found in Croft 1974.)

10.2 Stacked protection scheme and primitive operations

The Edman degradation will serve as our deprotection operation D.

To protect an aryl amine group Ar-NH_2 by an n-long polypeptide chain: we get a pre-made n-long polypeptide such as polyglycine. (Other aminos may be more suitable than glycine if one wants to make them solvent-philic or to enhance their resistance to certain reagents, e.g. nitrous acid.) Reacting this with $SOCl_2$ converts its carboxylic acid moiety to an acyl chloride, and then we react this with our aryl amine to protect it n levels deep. Degrading the peptide will eventually return us to a bare aryl amine. Because amines are powerful activators of aromatic rings for ortho and para electrophilic substitutions, while amides are only moderate activators, we will after full deprotection be able to accomplish appropriate reactions of the aromatic ring which we could not have done when the amine was protected.

[29] A systematic introduction to 1-level protection and deprotection schemes may be found in Carey & Sundberg 1990 (part B, pages 677-693); for a deeper discussion see Greene & Wuts 1991. An alternative to protecting and deprotecting a group, which is often good enough, although not in this instance, is instead to synthesize some precursor or "masked form" of that group, and then at the end transform it to the desired group.

For example[30], the following diazo coupling reaction hopefully will work when R- is H-, but not when R- is R''-CO-.

```
      HC-CH         HC-CH
      // \\     +  // \\       mildly acidic
R-HN-C     CH + NN-C     C-R'  --------------->
      \   /         \   /
      HC=CH         HC=CH
```

```
              HC-CH       HC-CH
              // \\       // \\
        R-HN-C     C-N=N-C     CH-R'
              \   /       \   /
              HC=CH       HC=CH
```

(Diazo couplings are preferentially para to the activator and only occur for strongly activated rings[31]. Ortho coupling could be prevented, if necessary, by methylation of the ortho positions.)

Consider the cases where (1) $-R'$ is an aldehyde (-CO-H) moiety or (2) is a halide -CHX-R. In case (2) the halide may be converted to a Wittig reagent by treatment with triphenylphosphine PPh_3 and a sodium alkoxide base (apparently no base is necessary if X is Fluorine), and the Wittig may be reacted with the aldehyde of the case (1). The result would be a joining of two arylamine "hands" via a bridge constructed of two aromatic rings linked by 3 double bonds conjugate to them, i.e. fairly rigid. (Note, it is possible to produce pure *trans* stilbene linkages [as shown] by controlling the reaction conditions or substituents, see Wheeler & B.de Pabon 1965.)

```
                         |   conjugated / aromatic   |     *********************************
                         | linkage between the hands |    **********************************
                         |                           |   ******             arm #2
                         |                           |  *****
                   HC-CH |     HC-CH       HC-CH     |  C-C
    hand #1        // \\       // \\  R    // \\       // \\
              H N-C     C-N=N-C     C-C=C-C     C-N=N-C     CH-NH      ----axis--
               2   \   /       \   /    H  \   /       \   /     2
                    C=C  |     HC=CH       HC=CH     | HC=CH          hand #2
                  *****  |                           |
    arm #1       ******  |                           |
*********************    |                           |
********************
```

Linkage of two arms with aryl amine hands, where the amines are both unprotected as peptide chains. Synthesized by the method described in the text. Preventing 180-degree rotation about the axis shown would hopefully be preventable by having appropriately bulky groups that are attached to the "arms."

Here all carboxylic acid groups would need to be protected before attempting the Wittig reaction[32]. After the completion of the joining reaction, the amines $-NH_2$ could be removed to prevent them from causing any trouble in

[30] I have been unable to find direct quantitative comparisons of the activating powers of these groups for diazo coupling in the literature. But experiments by Hashida et al. 1973 ("thin layer chromatography... showed the presence of only one azo compound in every case" whereas some small amount of a second compound would have been expected if the amide had non-negligible activating power) *suggest* the reactivity ratio is ¿100:1. Also in the literature (Zollinger 1995, Saunders & Allan 1985, Szele & Zollinger 1983) we find: "Aromatic amines (e.g. N,N dimethyl aniline) or phenolate ions [as activators] react very fast, in some cases close to diffusion [limits]." The reactivity order for activating power for diazo coupling is apparently Ar-O^- >Ar-NR_2 >Ar-NHR>Ar-NH_2 >Ar-OH$\approx$Ar-OR$\gg$Ar-Me>Ar-H. Ar-O^- is in fact 10^7 to 10^{10} times more reactive than Ar-OH. It has been speculated that the only reason that amides and related activators such as Ar-NH-SO_2, Ar-NH-CO-R, Ar-NH-CO-AR, Ar-NH-NO_2, and Ar-NH-CN can still cause diazo coupling is that the H attached to the N is made more acidic than in an amine, and if it deprotonates, the result will be an extremely highly activating negatively ionized N. Of course such deprotonation is very rare, especially in mildly acidic conditions. Both this fact, and the reactivity relation Ar-NHR>Ar-NH_2 suggest that better specificity could have been obtained if I had specified an amine of form Ar-NH-CH_3 (with corresponding protected form Ar-N(CH_3)-CO-R) instead of Ar-NH_2 in my tinkertoy design.

[31] Amine activation caused para:ortho product ratios ranging from 9:1 to 17:1 in table 6.14 p.284 Saunders & Allan 1985.

[32] The reactivity of Wittigs with amides is normally lower than with esters, which already are normally so far beneath ketones that they do not interfere. E.g. see P.J.Murphy & J.Brennan: J.Chem.Soc.Rev. 17 (1988) 1-. However, a few groups have succeeded in doing Wittig-Amide reactions, e.g. see J.Chem.Soc.Commun. (1981) 14-15; J.Org.Chem. 46 (1981) 2570-3; J.Org.Chem. 57 (1992) 3807-3814, usually by activation techniques or with relatively extreme conditions. Thus, this is a side reaction to keep in mind. For reviews of phosphorus ylide chemistry, see Johnson 1993 and Maercker 1965. Various degrees of "stabilization" of the Wittig by the use of various groups "R" may be tried in an attempt to reduce side reactions or improve stereoselectivity. Also, the Wittig reaction probably should be performed first and the diazo linking of the ylide side of the Wittig to the subtree, second. This might cut down on side reactions, but it would require some reaction redesign.

the future, e.g. by the $NaNO_2$ with H_3PO_2 (Cu_2O, pH 1) "permanent B kill" method of §13.

Thus, we have now described how to perform the deprotection operation D and the joining operation J. So we now have (at least on paper!) the *complete basic set* of synthetic tools needed in §7.1. We've assumed supplies of some suitable building block molecule are available in a form with protection by n-long polyglycine chains already appended. Since we have also described *how* to append n-level protection to amine groups, it seems highly likely that such pre-protected building blocks could indeed be synthesized (e.g. synthesize the "arms" each with appropriate protection appended to the arylamine group on its "hand," then join the arms by additional reactions at their shoulders).

It also seems possible to provide the more *advanced, optional set* of synthetic tools wanted in §7.1. The kill operation K could be provided simply by initially synthesizing building blocks with de-activated arms. For example, instead of an aryl amine (active arm) one could have an benzoic amide group. These groups could be unkilled (operation U) by the Hofmann degradation: Ar-CO-NH_2 $\rightarrow$Ar-NH_2 via the action of NaOBr. Finally, for a true kill – which cannot be unkilled – one could simply have nothing at all attached to that aromatic ring.

For global rigidizing (after unkilling and unprotecting to fully activate all aryl amine groups) one might be able to apply this reagent

```
          HC-CH
          // \\   +
  O=CH-C      CH-NN
          \    /
          HC=CH
```

followed by an intramolecular McMurry reaction (treatment with $TiCl_3$ and a zinc-copper couple; March 9-64 and McMurry 1989)[33] to convert any two nearby aldehydes (ketones also work well) into an alkene link:

```
                       TiCl
                           3
Ar-CH=O + O=CH-Ar ------->  Ar-CH=CH-Ar
                       Zn-Cu
```

The linkages produced will be almost exactly structurally identical to the ones produced by the previously described "operation J." Incidentally, the N=N double bonds in the azo links Ar-N=N-Ar in the J linkage should come out *trans* because these bonds have a low rotation barrier and the *trans* form is sterically favored[34]. Again this steric favoritism could be increased if necessary by the intentional adjoining of bulky groups.

There are also possible alternatives to the plans we've just suggested.

One is to base the joining reaction on some different chemical capability that aryl amines have, but regular amines (and peptides) do not. One possibility is the conversion of an aryl amine (Ar-NH_2) to a diazonium salt (Ar-$N^+ \equiv$N) with nitrous acid HONO (actually $NaNO_2$ with an appropriate acid, e.g. HCl). Because diazonium salts are synthetically so versatile, this could be followed by any number of manipulations, perhaps including a conversion of the diazonium salt to an aryl hydrazine (Ar-NH-NH_2) with sodium sulfite Na_2SO_3 followed by acidic H_2O, and then this aryl hydrazine could be used in a Fischer indole synthesis as in §9.2 and §13. Here, one would have to rely on the different vulnerabilities of aryl and aliphatic amines to HONO: HONO attacks aryl amines at pH 1, while[35] aliphatic amines such as those in the peptide protective chains are not attacked unless the pH is ≥ 3 (March 2-49, Kornblum & Ifland 1949).

Another idea: Instead of using the difference in *activation* of an aromatic ring arising from the attachment of an amine instead of an amide as the basis of our protection scheme, we could instead use the *steric* effect of the protective chain, by finding some reaction which is prevented if the chain has length ≥ 1, but which can occur if the chain does not exist. Nor do the protecting chains need to be polypeptides. The same idea (multi level protection based on steric effect) could be used with any other kind of chain. All we need is the ability to synthesize and adjoin an n-long chain initially, and the ability to degrade exactly one monomer off the end of each chain. For example, the "Ruff degradation" (§34.9 of Morrison & Boyd 1992), removes one monomer from the aldehyde end of an aldose chain H-$(HCOH)_n$-CHO. The inverse of the Ruff degradation is the "Kiliani-Fischer synthesis" for adjoining an HCOH unit.

[33] A side reaction is also possible with amides here. McMurry 1989: "Amides react slowly with low valent titanium but are compatible with carbonyl coupling unless extended reaction periods are used."

[34] Saunders & Allen 1985 remark that the conversion from the *cis* [also called "labile," "syn"] form to the *trans* [also called "anti," "iso," or "stable"] of azo linkages is "spontaneous, often rapid," and gives rate constants for this change in table 5.1 page 213 ranging from $(.5\text{-}2)\times10^{-3}$/sec for diazo-sulfonate salts. A later book on the same subject is Zollinger 1995

[35] Conditions would also have to be adjusted to minimize any hydrolysis of the amide links of the peptide chains by the HONO.

11 A tinkertoy design based on DNA hybridization

The plan of §7.2 is conceptually so straightforward that one must only present a design for the tetrahedral "hub" and say how it is to be attached to its 4 double stranded DNA "arms." Or more simply, as we've argued in §5, only *three* arms are really required.

I will not design such a hub molecule here, but I will try to suggest some chemistry capable of performing the key arm-hub attachment operations.

To perform arm-hub attachment requires 6 sites to be available on each hub at the right separations and angles (since each of 3 arms is double stranded, and each strand needs to be attached). At first the requirement for 6 different joining reactions, each mutually compatible and compatible with DNA, seems impossible to handle. But we will see how to reduce this requirement down to just one reaction.

The first idea will be to link each arm to a "shoulder" which has an attachment point for each of the arm's two DNA strands. As we've previously argued (§7.2) these two attachments need only be single bonds[36]. Then the 3 shoulders are to be linked to a "core" to form the full hub-arm assembly. This core may be assumed initially to be attached to one of the shoulders, and the only problem is then to attach the two remaining shoulders. This may be accomplished by only one type of chemical reaction, if we assume that the core is provided in a form in which one of the two shoulder attachment sites is initially "protected" by a removable group. The core-shoulder links should involve ≥ 2 bonds to assure rigidity.

Now during the usual mechanized solid-support based synthetic processes for making DNA, the "shoulder" could have been joined to the solid support originally by a "fusible link" and the growing DNA strand initially affixed to it instead of directly to the solid support.

Also, some special group "S" could be adjoined to the other end of a growing DNA strand, by the same chemistry the DNA synthesizer uses to adjoin a new DNA monomer.

Thus we only need to devise one new chemical reaction, for use in joining S to the shoulder molecule to get a shoulder with two complementary and reverse direction DNA single strands attached. Because the design of S and the shoulder are completely under our control, this should be possible despite the mild conditions required to stay DNA-compatible. The *same* reaction could then in fact be used to accomplish the core-shoulder attachments, provided the shoulders were initially provided with their core-attachment sites protected.

So in summary, we've now reduced the problem to devising one new joining reaction and two kinds of deprotections, all of which have to be DNA compatible.

A natural possibility for the joining reaction is then the Robinson annulation (cf. §9.2) with protection of the ketone groups on each side as suitably chosen ketals to allow deprotection with sufficiently mild acid.

```
R       Apply protection with              R   O
 \      HO-CH -CH -OH & acid catalyst,      \ / \
  C=O       2   2    dehydratory conds.      C   CH
 /      ---------------------------->       / \  | 2
R                                          R   O-CH
      <-----------------------------------         2
        Remove protection with acid hydrolysis

Protection of ketone group as cyclic ketal. Protection may also be applied by reacting
with the disilyl ether (CH ) Si-O-CH -CH -O-Si(CH )  .
                          3 3      2   2       3 3
```

Note that DNA contains neither ketone groups nor non-aromatic C=C bonds, hence should not interfere with the annulation reaction. Note also that ketals will be stable under the basic and dehydratory conditions required during the Robinson annulations.

12 An extremely simple planar tinkertoy design based on directed trees

It is very easy to devise designs based on the directed tree concept of §7.4. For example, we'll now show that the peptide synthesis and protection chemistry we've already discussed (§10) suffices to implement "operation set #1" of §7.4, but most chemists should have no trouble devising similar plans based on other chemistry.

Here is a complete design for a planar tinkertoy molecule.

[36] the "shoulder" can look like a fake DNA monomer unit, and hence the attachment will lead to a rigidity just as great as that enjoyed by the normal attachment between the left and right halves of some DNA

```
            CH      Top ("T"):
            | 3   Tert-butyl ester
          O-C-CH    protection for
         /  |   3    carboxylic acid.
      O=C   CH                                            H
       |      3                                           C    O
        C              Cl                                / \\  ||
       / \\            |          Leg B:               HC    C-C
    *HC    CH*         C     2,4 dichloro benzyl        ||   |  \
      ||   |          / \\     carbamate               HC    C   O
      C    C   O    HC    CH   protection                \ // \ /
     / \ // \  ||     ||   |     for amine.               C    C
  H N   C    N-C   C-C    C-Cl                            H    ||
   2    H    H  \ /   \ //                                     O
                 O     C                              Phthalic anhydride,
Leg A: amine.          H                         mentioned below as possible
                                                replacement for carboxylic acid
   Possible Tinkertoy for "directed tree,"    moeity at top. In a "phthalimide,"
       based on peptide chemistry.           the rightmost O is replaced by N-R.
```

The two legs A,B of the tinkertoy are both amines (one provided protected as a phthalimide or a 2,4 dichloro benzyl carbamate, however) and the top T is a carboxylic acid (but provided protected as a tert-butyl ester).

"Deprotect T" by acid hydrolysis. "Join A" by activating the deprotected-T acid from the subtree with $SOCl_2$ and then reacting with the amine. Then B's phthalimide protection may be removed by hydrazine or $NaBH_4$ treatment (see below in §13) – or if it is protected as a 2,4 dichloro benzyl carbamate, by Pd catalyzed hydrogenation – then joined the same way as A was. Since the resulting amide links have been designed to form conjugated to aromatic rings (same structure as in Kevlar, cf. §8.1.2), the result will be reasonably rigid and planar.

We must prevent the subtrees from linking onto the amine legs twisted 180° "the wrong way." Actually there are two ways to go wrong stereochemically. First, the peptide bond could come out "cis" instead of "trans" (compare the two thioamide atropoisomers pictured near the beginning of §8.1.1). The undesired "cis" peptide conformation is highly energetically disfavored because hydrogens (at the *'d positions) on the two benzene rings would overlap. Second, the entire subtree being bonded to one of the legs of our tinkertoy could "flipp over" 180°. In fact this could happen to both subtrees. It seems to me that by adding bulky groups to the tinkertoy design, that only 1 of the 4 possible flips of the two subtrees would be sterically favorable. For example, one of the two *'d H's (two flavors of tinkertoys would be needed) could be replaced by tert-Butyl groups -$C(CH_3)_3$. ("Endcap" legless tinkertoys would also need to be available.)

For more rigidity with similar chemistry, one could use phthalimide linkages (resulting in more rigidity than our amide linkages) gotten by the reactions of phthalic anhydrides and primary amines. The tinkertoy design would be essentially the same but the protected acid at the top would have to be replaced by a protected phthalic anhydride group.

It should be possible to try to implement a "global rigidizing" operation, the best way seems to be to design special "end cap" tinkertoys which can later be "activated" to support such a reaction.

13 A highly rigid planar tinkertoy design based on directed trees

Since we've just seen how easy it is to design a directed-tree schema tinkertoy, in this section we'll undertake a more difficult and rewarding task: a design for a planar, *highly rigid* tinkertoy in which all joining reactions are Fischer indole syntheses, to create highly rigid fully aromatic fused ring linkages.

The tinkertoy molecule is shown below[37]. It is based on triphenylene, but once our reactions are revealed, the reader should have no trouble devising similar tinkertoys that are essentially the same but are instead based on either copper phthalocyanine[38], triptycene, or 1,3,5,7 tetraphenyl adamantane[39].

[37] I do not know how to synthesize it. If you want to try to plan a synthesis, begin by consulting Buess & Lawson 1960 and Robinson 1982.

[38] Which would permit 4-way instead of 3-way branching to create subgraphs of the square lattice instead of the hexagon tesselation. Some additional reactions would be needed to provide the increased selectivity; it would be a good exercise for the reader to try to devise some. (Hint: consider aryl nitro groups.)

[39] Which would permit 4-way tetrahedral branching. These molecular structures have been presented in §9.1.

```
                 H
(i)              C
                / \\               (ii)
             *HC    C-I
      H   H    |    |           <-----rootwards
  H   N   C    C    C-CH
   \ / \ / \\ / \\ /    3              A
  H-C   C    C    C                   /
    |  ||    |    |                T-*
  O=C---C    C    C                   \
         \ // \ // \                   B
          C    C    CH
          H    |   ||           Rigid planar tinkertoy.
               C    CH      (i) molecular structure.
              / \\ /        (ii) concept.
           H C    C
            3     |
                  NH
                    2
```

Actually, there is a slight problem with the formula above – it probably will not exist because the ketone group on the left will prefer[40] to tautomerize to the more aromatic enol form! (The resulting structure is called an "indoxyl.")

```
      H   H                       H   H
  H   N   C                   H   N   C
   \ / \ / \\                  \ / \ / \\
  H-C   C    CH                 C   C    CH
    |  ||    |                 || ||     |
  O=C---C    CH              HO-C---C    CH
         \ //                        \ //
          C                           C
          H                           H

  unstable ketone form        indoxyl (enol form)
```

However, the ketone form above is useful mentally. *Actually* we shall assume that the molecule is provided to us in a form with a *protected* ketone, which ought to be able to exist stably. There are two ways to protect ketones $R_2C{=}O$ that seem plausible. The first way is as a cyclic dithioketal as in Carey & Sundberg §13.1.3, March p. 375, or Greene & Wuts 1991.

```
R       Apply protection with                R   S           (Possible "catalysts" [all
 \      HS-CH -CH -SH & catalyst              \ / \           Lewis acids] include:
  C=O        2   2                             C   CH
 /      ---------------------------->         / \  | 2     ZnCl , BF , Mg(O SCF ) , Zn(O SCF ) . )
R                                            R   S-CH          2    3      3   3 2      3   3 2
      <-----------------------------------           2
        Remove protection with Hg++ salts

Protection of ketone group as dithioketal. Protection may also be applied by reacting
with the dithiosilyl ether (CH ) Si-S-CH -CH -S-Si(CH )  with catalytic ZnI .
                              3 3       2   2        3 3                   2
```

Actually we will only have need for the deprotection reaction. Everything else we discuss will be performed with T protected[41].

The second way is as a dimethyl alkene $R_2C{=}C(CH_3)_2$, in which case the protection may be removed by cleaving the C=C bond in the cold with ozone O_3 followed by Zn/H_2O workup. (Also, this kind of protection of ketones

[40] Actually, even if it did exist stably, there would still be an annoying side reaction: ketones $R_2C{=}O$ can react with amines $R\text{-}NH_2$ to form imines $R\text{-}N{=}CR_2$ (this happens best in hot, acidic, and dehydratory conditions) so the tinkertoy, as shown, might self-react. This is prevented if the ketone is assumed to be provided in protected form.

[41] Other possible ketone-precursor forms include: nitro groups $R_2CH\text{-}NO_2$ may be converted to ketones $R_2C{=}O$ with aqueous $TiCl_3$ or various other modernized "Nef reactions" (see March's reaction 6-4); and vinylic halides $R^1R^2C{=}CXR^3$ may be hydrolysed to ketones $R^1R^2CH\text{-}CO\text{-}R^3$ with mercuric trifluoroacetate (March 0-1). But these two kinds of groups seem too vulnerable to the sorts of reagents we will want to employ later.

could be applied with a Wittig reagent $Ph_3P{=}C(CH_3)_2$, but again we only need the deprotection reaction.)[42]

We will now show how to implement "operation set #3" of §7.4.

First, "joining B" may be accomplished by converting the amine to a diazonium salt with nitrous acid[43] HONO, and then to a hydrazine with sodium sulfite Na_2SO_3 followed by acidic H_2O,

```
              HONO        +     1. sodium sulfite
Ar-NH       ------>  Ar-NN   ------------------->  Ar-NH-NH
     2                          2. acid workup               2

an aniline     a diazonium salt             an aryl hydrazine
```

and then reacting with the deprotected ketone moiety[44] of the subtree we wish to join on according to the Fischer indole synthesis as in §9.2.

Note that this will create an "indoloindole" linkage

```
            H       H   H
            C       N   C
          // \     / \ / \\
         C    C---C   C    C
   ...   |    ||  ||  ||   |   ...      Indolo-indole linkage.
         C    C   C---C    C
          \\ / \ /     \ //
            C   N       C
            H   H       H
```

in which any bending due to the downward pointing pyrrole pentagon, is exactly compensated for by the bending in the opposite direction due to the upward pointing one, so the linkage is straight.

Third, once B has been joined (or otherwise deactivated), we may "activate A" by converting the iodine (or a different halide atom could be used, if it works better) to an amine either by use of $NaNH_2$[45], or by a "Gabriel synthesis" with potassium phthalimide (with Cu_2O or CuI catalyst) followed by a hydrolysis or hydrazine[46] treatment (March 0-58). Then the "join A" operation may be completed exactly similarly to the just described "join B" operation.

We thus have the *complete basic set #3* of operations required in §7.4 to synthesize any spanning tree of any subgraph of the periodic regular hexagon tiling in the plane. The molecules we synthesize will be rigid, planar, and fully aromatic (and hence in fact should be electrically conductive).

It is also possible to perform some *more advanced, optional* operations.

---

[42] Note, the choice of a dimethyl alkene was intentional since these seem to be the most vulnerable to O_3 (Razumovskii & Zoukov 1984, see esp. tables 3.12 p.115 and 9.6 p.359) with a rate constant of 5×10^5 in Razumovskii's units, and in our case, maybe even higher due to the aromaticizing enolization providing additional drive. Undesirable side reactions with ozone can include: cleavage of C=N and N=N bonds and aromatic rings (but the rate constants for these things are only .3-15 per moderate sized aromatic molecule); reactions with C-C bonds (< 1), and oxidation of amines. Although amines of all varieties oxidize at fairly high rates (2×10^3-10^8), the molecules we plan to deprotect will only have N's incorporated inside aromatic heterocycles such as in quinoline, whose rate constant is only 4.2. Other side reactions of ozone, not relevant to us, include reactions with alkynes (60-120), C=S bonds (2×10^6), nitro groups, phenols (200-9000), and many sulfur compounds.

[43] A side reaction which must be avoided is: the HONO must not remove the ketone's protection. This might be a problem if dithioketal protection is used, because El-Wassimy et al. 1983 note that 3 equivalents of $NaNO_2$ in 4 molar aqueous HCl for .25-20 hours, depending on the dithioketal, will remove dithioketal ketone protection.

Another side reaction, if we are instead using dimethyl alkene protection, is Meerwein arylation of alkenes by aryl diazonium salts (March 4-19). This is only suposed to happen if the alkene is "activated" by an electron withdrawing group (which it isn't) and if there is a copper chloride catalyst (which there isn't), but this might still happen to some degree.

[44] The Fischer indole synthesis can still proceed despite aromatic enolization of this ketone, e.g. see Robinson 1982 page 43 for an example with napthol. However, according to Robinson p.356, nobody has yet investigated to see whether the Fischer indole synthesis will still proceed in this particular kind of case to make the particular 6-5-5-6 fused ring structure (indoloindole) we are trying to manufacture. Indoloindoles have been made by other means, however. In the event that indoloindolization does not work, various modifications of our planned linkage to make symmetric carbazoles instead (see supplementary figure 1) should still work. In the event that our preferentially-enol "ketone" is regarded as a big handicap, one could make a genuine ketone, but this would require (if we want pure fused aromatic structure) additional aromatization steps to be added to the synthetic procedure such as in the "Borsche carbazole synthesis" (Robinson 1982).

[45] If the *'d H in the tinkertoy formula is replaced with an inert electron withdrawing group such as $-O-CH_3$, then the action of $NaNH_2$ on the aryl halide will be less likely to cause cine substitution (motion of the adjoined $-NH_2$ group to an adjacent position on the benzene ring, caused by a benyzne intermediate), according to Fieser & Fieser volume 4, page 439. $NaNH_2$ unfortunately will react with $-NO_2$ and vinylic halides, suggested above as possible ketone precursors for optional use in place of dithioketal ketone protection.

[46] If our ketone is protected as a dimethyl alkene, then a possible side reaction would be hydrogenation of the alkene by the hydrazine, see March p.774-779. However, this only happens in the presence of oxidizing agents such as O_2, H_2O_2, or cupric ion, which may be excluded. Also, March 0-58 reports several alternatives to the use of hydrazine in the Gabriel synthesis, namely Na_2S in aqueous THF or acetone, $NaBH_4$/iso-propanol followed by acetic acid, 40% aqueous methylamine, or n-pentylamine.

T may be permanently deactivated by various methods for removing ketone functionalities $R_2C{=}O \rightarrow R_2CH_2$, such as (if we are using dithioketal protection) catalyzed hydrogenation of the dithioketal, (or if we are using dimethyl alkene protection) addition to the alkene to deactivate it, or reduction of the ketone/phenol, or ketones may be converted to difluorides with SF_4/HF (Boswell et al. 1974).

B may be *protected* and *unprotected* by several methods in Greene & Wuts 1991, e.g. protection as a phthalimide via phthalic anhydride[47] with deprotection by hydrolysis or hydrazine treatment as above, or alternatively again with benzyl chloroformate, with deprotection ultimately by H_2/Raney nickel or hydrolysis in aqueous acid below pH 1. That could serve as a "kill and unkill" operation. An *irreversible kill* of B could be accomplished by the action of $NaNO_2$ in hypophosphorous acid H_3PO_2, with catalytic Cu_2O, which performs the conversion Ar-NH_2 $\rightarrow$Ar-H through a diazonium salt intermediate. More simply, one could always have an alternate kind of tinkertoy available without any attached -NH_2.

Similarly *A* may be *irreversibly killed* by reduction of the aryl iodide Ar-I$\rightarrow$Ar-H by a metal hydride such as $NaBH_4$ (with catalyst), NaH, Ph_3SnH (see March 1-42), or Mg in isopropanol (Carey & Sundberg reaction 1 scheme 5.9 p.259), or by replacing of the -I by a short alkyl group -R using an organoborane with Pd catalyst (Carey & Sundberg reaction 6 scheme 9.6 page 464).

Again, a simpler way is just to have an alternate kind of tinkertoy available with no attached iodine.

A more interesting way to kill Ar-I groups is to do an organopalladium "Heck reaction" (March 4-20; Carey & Sundberg 8.2 p.418; Heck 1982; Tsuji 1980; Patel et al. 1977) or a reaction with an allylic Nickel complex π(2-methylallyl)NiBr (March 3-16; Carey & Sundberg 8.3; Semmelhack 1972). to convert them to an olefin Ar-CH=CH_3.

A "global rigidizing" operation G at the end of the spanning tree synthesis (thus "inking in" all the bonds of the planar hexagon tesselation that were not part of the spanning tree we made) could then be accomplished as follows.

Assume we'd previously "killed" all the A's by converting them to appended alkenes as we said, and all the B's by protecting them. We now unkill all the B amine groups, and induce links between the alkenes and the amines from the B legs in a couple of possible ways:

1. One way is to convert the olefins to bromides with Br_2 and then the phenols and bromides may be linked by an intramolecular Williamson ether synthesis (March 0-12) by the action of KOH in dimethyl sulfoxide (or HgO and HBF_4 in CH_2Cl_2).

2. Another way is to diazotize the Ar-NH_2's and then convert them to phenols Ar-OH by adding Cu_2O to an aqueous solution with a large excess of $Cu(NO_3)_2$ (March 3-20), and then the phenol may be added to the olefin to forge an ether link (March 5-4), catalyzed by H_2SO_4.

```
                         acid catalyzed addition reaction
Ar-OH + CH =CH-R   -----------------------------------> Ar-O-CH - CH - R
          2                                                    2    2

   Formation of ether link as in "global rigidizing" operation.
   The addition shown here is anti-Markovnikov, which would presumably
   be forced by geometric length constraints despite the fact that
   this addition is normally Markovnikov. In any event, an additional
   methyl group in place of one of the Hs on the lefthand CH , would
   make the desired addition OK under Markovnikov's rule.   2
```

In order for this to work, it would have to have been arranged that every non-tree (but desired) bond in the hexagon tesselation had -I's at one end and -NH_2's at the other. However, if a reflected tinkertoy with switched A and B leg types is also available, this is no problem[48].

14 Cleavage reactions

Alkenes $R^1R^2C{=}CR^3R^4$ may be cleaved to ketones $R^1R^2C{=}O$, $O{=}CR^3R^4$ by the use of ozone O_3 at low temperature to form an unstable ozonide, followed by treatment with H_2O in the presence of zinc. (Aromatic "double bonds" [and endocyclic double bonds generally] are attacked less readily but still can often be cleaved.)

Esters may be cleaved to acids and alcohols, and ethers to alcohols, by hydrolysis; and amides may be cleaved to acids and amines by ammonolysis (NH_3 in alcohol).

[47] Any reaction of this phthalimide with hydrazine moieties as formed above for the Fischer indole synthesis, would have to be avoided. This is probably achieveable since probably the Fischer indole reaction will probably be a much faster reaction.

[48] A hexagon tesselation with a spanning tree removed is a graph consisting entirely of vertices with degree$\leq$ 2, hence consisting of paths and cycles only. By arbitrarily assigning directions to all the cycles or paths formed by the desired nontree bonds, we see that a solution always exists.

Nonenolizable ketones may be cleaved with $NaNH_2$, see March's reactions 2-45 and 2-46.

Azo links Ar-N=N-Ar may be cleaved by stannous chloride $SnCl_2$.

Aryl sulfones Ar-SO_2-R are cleaved by $NaNH_2$ in piperidine (C_5NH_5) to aryl piperidines and sodium R sulfites, with emission of ammonia.

Disulphide links R-S-S-R′ may be cleaved using cold performic acid (H-CO-OOH) to sulphates R-SO_3H and R′-SO_3H.

The highly toxic reagent BrCN (Cyanogen bromide) cleaves polypeptide chains on the carboxyl side of methionine residues only (a reaction discovered by E.Gross & B.Witkop and used in protein sequencing):

```
                                BrCN
R-CO-NH-CH-CO-NH-R' ------------>   R-CO-NH-CH-CO   + NH -R'
        \            in F C-COOH            |  |        2
  H C      CH              3              H C  O
   3 \    /  2          (12 hrs)            2 \/
      S-CH           or 70% formic acid        CH
          2          & 30% water (24 hrs)        2
```

BrCN will also cleave thioethers R-S-R.

All these reactions are possible candidates for use as operation C, "cutting fusible links," but unfortunately none of them seem very hopeful for being chemically compatible with all our previous schemes[49].

15 Applications

I would expect the following fairly immediate applications.

1. "Photonic bandgap" materials. Analogues of these materials have long been known in one dimension as optical multilayer mirror coatings. However, the creation of a genuine 3-dimensional dielectric material with a photonic bandgap (that is, a range of wavelengths throughout which light propagation is impossible) has not been accomplished so far (except at macroscopic wavelengths), despite considerable effort. It is known (Joannopoulos et al 1995 and references therein) that any material whose refractive index is any of several known periodic functions (with period of the same order as the bandgap wavelength) will do the job, the problem is to create that material. By adjoining groups of high polarizability (or not) to our building blocks we will have complete control over the refractive index, and it ought to be possible to synthesize a "unit cell" macromolecule roughly a 500nm cube in size and shape. In particular, large aromatic regions exemplify "high polarizability," while nonconductive low density regions are the opposite. (Especially high synthetic reliability, i.e. getting every atom right in every molecule, may not be required in this application.) By crystallization of this macromolecule, a photonic bandgap material will result.

   Unfortunately, the known theoretical constructions of 3D photonic bandgap materials require a refractive index ratio $\eta_{\max}/\eta_{,\min} \geq 2.0$ (based on spherical cavities arranged with the symmetry of a diamond lattice) or ≥ 2.1 (arrays of drilled cylindrical holes), while most organic compounds only have refractive indices in the range 1.25-1.75.

   The lowest index possible is $\eta = 1$ (vacuum). Perfluorinated nonaromatic substances such as C_6F_{14} tend to have the lowest indices (1.2515) among dense packed organics, and the largest indices are from large aromatic molecules. For example, in the sequence 1-Chloro(Benzene, Napthalene, Anthracene), the refractive index η increases monotonically (1.5241, 1.6326, 1.6959). The pigment quinacridone red, with 5 fused aromatic rings in a row, has $\eta > 2$.

```
                    O
    H    H   H   ||   H
    C    N   C    C   C
   / \\ / \ / \\ / \ / \\
 HC    C   C    C   C    CH*
  ||   |   ||   |   ||   |
*HC    C   C    C   C    CH
   \ // \ / \ // \ / \ //
    C    C   C    N   C
    H    ||  H    H   H
         O
```

Quinacridone red.
"Magenta" has methyl groups attached instead of H's at the *'d carbons.

[49]The last few seem the most likely to work, but $NaNH_2$ and BrCN are very powerful reagents likely to have too many side effects, and disulphide links tend to be fragile – they are also cleaved by too many other reagents one might want to use elsewhere.

However, if we assume that in the future, theoretical constructions of photonic bandgap materials will be found which require smaller index ratios (and this seems likely, since the present constructions are geometrically very simple), and if we also assume that we will be able to link large aromatics to some tinkertoys and not others according to our whim, then these materials will become possible.

2. Extremely low density materials. "Aerogels" are made from alcoholic or aqueous silica gels by various ingenious methods (based on "supercritical fluids") that aim to remove the solvent without creating the surface tension forces that would normally collapse the gel network[50]. The result is "solid smoke," a very low density randomized network of silica strands. It seems to me, however, that our methods would enable replacement of the silica by organic molecules (saving mass) replacement of random structure by precisely specified structure, and replacement of flexible chains by rigid bars, increasing strength and therefore also allowing smaller density. We could presumably also introduce other controllable properties by modifying our building block molecules (e.g. the adjoining of arbitrary chemical groups, perhaps electrical conductivity, etc.). Finally, it is believed that the structures with the highest ratios of compressive strength to weight are in fact fractal in structure. (Preventing the buckling of large beams is accomplished with bracings of angled smaller beams, and preventing buckling of these is done with even smaller beams, and so on. See a picture of the Eiffel tower.) Creating such fractals should be no problem with tinkertoys, with the resulting material of nonzero compressive strength in fact having, ultimately (for a fractal dimension < 3), density *zero*.

3. Creation of "molecular sieve" materials with arrays of precisely spaced and sized holes. Certain "zeolite" minerals already do this, but now we will be able to control the hole sizes and spacings, and also to control which chemical groups appear where inside the holes.

4. Creation of micromolds and/or microstencils to allow the creation of smaller metal, semiconductor, etc. structures than can currently be made.

5. The beginnings of research into "artificial enzymes" and nanoscale mechanical and electrical mechanisms will now become possible.

Really, though, it is impossible to predict the applications that might happen once a tinkertoy synthesis machine became widely available.

Of the above list, I think only the latter 2 (microstencils and research) applications will be of especial importance, because the materials applications will probably be too expensive for large scale production. This will be due to the large number of synthetic steps required and the expensive reagents; the yields may not be all that bad. Recall that in Merrifield's linear synthesis of ribonuclease (a 124-amino polypeptide, synthesized using a total of 369 reactions and 11931 mechanical operations) the final yield was 18%, hence the yield per amino-addition was $.18^{1/124} = 98.6\%$. So it seems plausible that after enough optimization of tinkertoys and tinkertoy processes has been done, yields of 95% per tree-bond joining will be achieved. In that case, synthesizing a 10^3-tinkertoy molecule as a fairly well balanced binary tree would require synthetic depth $2 \times \log_2 10^3 \approx 20$ (the factor 2 is since two joining reactions are needed in our schemes for doubling the tree size) for a total yield of $.95^{20} = 36\%$. A 10^6 tinkertoy molecule (mass$\approx 10^{10}$ Daltons, far larger than any molecule yet synthesized reproducibly) would have yield $.95^{40} = 13\%$.

In the longer term, it might be possible to make smaller and faster computers with the aid of molecular tinkertoys. Several paper designs for molecular-size computer components, such as "wires," "diodes," and memory bits that work by altering the valency of an atom (for example see Robinson & Seeman 1987 and references therein) have been proposed. It seems to me that the biggest obstacle between such designs and reality is the problem of assembling the structures they are talking about so that experiments can be carried out on them. I do not think that any "self-assembly" process is going to be good enough to assemble substantial circuits. But our tinkertoy proposal has that potential.

16 "Eutactic Nanosystems"?

What about the "nanotechnology revolution" visions enunciated in K.Eric Drexler's books? (Drexler 1992 is the most scientific of these works. Popular treatments include Drexler et al. 1986, 1991 and Regis 1995; a brief sketch of nano religion is Amato 1991.) In my opinion, molecular tinkertoys still will not suffice to accomplish those visions. To see why, let's discuss and critique Drexler's ideas a little.

First, I should say that a large amount of hype and pseudoscience has attached itself to nanosystems. Upon encountering that, a skeptical reader such as myself becomes extremely tempted simply to dismiss it immediately

[50] Similarly gentle methods might be required during solvent removal operations to prevent the collapse of our tree structures at intermediate stages of our synthesis.

as silly. However, it is better to try to peel away the hype and try to ask what genuine content there is and what is genuinely either scientifically conceivable or technologically possible.

Drexler has coined the word "eutactic" to denote a new state of matter in which *every* atom is in a precisely specified location. He imagines macroscopic amounts (e.g. a kilogram) of eutactic matter with molecular structure rather like an intricately carved perfect crystal of diamond, shaped into complicated mechanical structures resembling Charles Babbage's mechanical computers reduced to the nanometer dimensions. Assuming (and this is a big assumption, which neither Drexler nor anybody else have come close to realizing) such things could somehow be created at will, we would then presumably be able to make immensely fast and parallel computers with enormous memories and incredibly low power requirements. (Drexler 1992 in his "introduction and overview" claims that 10^{16} instructions per second per watt "appears to be both physically possible and practically realizable," and certainly thinks random access memories with 10^{19} bits should be ditto.) Also, we could make self reproducing "von Neumann machines" which could convert raw materials (i.e. non-eutactic matter) and supplied power into armies of additional copies either of themselves or of other amazing eutactic nanomechanisms (such as micron sized robots). With these nano-robot armies at our command, it would then be no problem to produce large amounts of immensely strong and chemically inert material (exceeding diamond) in arbitrary forms, revolutionize medicine, computing, and manufacturing, and generally achieve utopia. A chemist's immediate gut reaction would be that the enormous surface area of such structures would make them extremely vulnerable. But Drexler's claim is that nothing would ever go wrong and the nanomechanisms would never wear out because their atoms are locked in place and available thermal energies, frictional forces, etc. won't ever be sufficient to knock even one atom out of place. Also, chemical degradation won't ever happen because eutactic matter doesn't even have one atom of any chemical inside it to degrade anything! Cosmic rays could be a problem, but Drexler claims that with known mathematical techniques for making reliable mechanisms from unreliable components (these techniques are certainly described highly inadequately in Drexler's books), this problem is surmountable. Of course degradation could occur at the boundary between the eutactic and noneutactic worlds, but one could use thick shields to keep out the crud[51].

All this sounds great, but if nobody can ever make enough eutactic matter to get the Drexler vision started, it will never get started.

There are already a large number of self reproducing nanoscale mechanisms around, known as "life." However, life as we know it involves eutactic matter nowhere (or hardly anywhere!), and thus seems far inferior in principle in its capabilities to Drexler's visions. Question #1: If Drexlerian eutacticity is so wonderful, why did evolution and life (which certainly have developed the art of nanoscale manipulations to a much higher plane than any current human technology, and which have been experimenting with them for billions of years on a massive scale) never create it? Presumably because it is very difficult to create, or because it doesn't work. And even if it does work... question #2: how does Drexler expect that human technology is going to be able to produce large amounts of eutactic matter, when human technology is currently incapable either of rivaling, or of creating, life, which as we've just said is apparently *still* not sufficient to get to where Drexler has to go?

And here is a final question to ponder. Drexler's self-reproducing nanobeasts have to convert the ugly non-eutactic world into eutactic matter in order to reproduce and generally to do "molecular manufacturing." In order to do this, the eutactic world has to come into contact with the non-eutactic world, which automatically pollutes it and thereby instantly destroys eutacticity. To get around this, Drexler envisions "molecular mills" which consist of a cascade of n stages of various nanomechanisms intended to perform chemical purifications and atomic motions, at each stage making incoming raw materials a constant factor purer and more precisely located than they were at the previous stage. Question #3: in the initial, impure and non-eutactic stages of molecular milling, why are the nanomechanisms going to keep on working without wearing out and getting degraded and clogged by the continual assault of free superoxide radicals and general gunk in the outside world? Drexler's answer would seem to be that one has to keep replacing and repairing those mechanisms faster than they degrade – which is the same strategy that life adopts everywhere – but then in what ways do Drexlerian nanotechnology robots really remain superior to the usual kind of non-eutactic life? It seems to me that the main ingredient of life is the constant self-repair and fault tolerant capabilities, hence these should also be the main secret of nano-manufacturing, but Drexler hardly discusses these things at all. And if they are not superior, why bother with nanotechnology, why not just do genetic engineering?

Even assuming there aren't any satisfactory answers to these questions, it still seems clear that any future technology giving us as much nanoscale control as life currently has, would permit great improvement in many areas

[51] Drexler 1992 constantly makes various highly detailed designs for nano-parts, e.g. figure 10.17 is a "2808-atom sleeve bearing," figure 10.32 is a 3557-atom planetary gear system. Drexler's claims that these would work are based on molecular modeling and simulations. The problem with that is simple: there currently is no method for modeling large molecules and their interactions which can be trusted. In areas such as this where computational proofs are infeasible, one must resort to experimental proof. I have no objection to Drexler trying his best to model these things; I just want to reject the notion that this has proven anything. It's not a proof, it is a plausibility argument. (Drexler 1992) is a large number of plausibility arguments piled on top of each other. That is also OK, provided one recognizes this, but reading Drexler's introduction, one may not recognize this.

of current technology, including computers and materials.

And on the other hand, if we totally buy the validity of the eutactic nanosystems vision, we are still left with the narrower question of whether the present paper's "molecular tinkertoys" would enable "eutactic nanosystems." Although they are certainly going in the right direction, as far as I can see, the answer is no.

For one thing, Drexler inherently seems to require the very high rigidity and chemical inertness of thick chunks of diamond lattice. The structures one can make with molecular tinkertoys (and every other known precisely controllable technique from synthetic organic chemistry, for that matter) are considerably less rigid and inert.

More importantly, even though molecular tinkertoys should be able to produce decent yields of macromolecules with precisely specified and reproducible atomic structure, this is still a long way from being eutactic matter. How can we perfectly purify the product, removing *every* molecule of solvent, leftover reagents, impurities, metal atoms adhering by van der Waals forces, etc. from our product? This seems to be higher standard of chemical purity than is generally achievable, and even in the few cases where maybe it is achievable, harsh conditions tend to be required, which would probably be injurious to our tinkertoy constructions[52].

17 Conclusion

Chemistry used to be difficult and mysterious. Now (assuming the ideas of the present paper get implemented), a large subset of chemistry becomes a triviality. You just draw the shape you want on your computer screen and it does the rest of the mental and physical work.

Chemists have devoted great ingenuity to devising syntheses of molecules of various strange shapes, for example "catenanes" (linked loops), "rotoxanes" (knots), Moebius strips, tunnels, and hollow shells. Once the tinkertoy building machine is available, no more ingenuity will be needed; the creation of these things and more (e.g. gearwheels, "universal joints," Peaucellier linkages for converting straight line motion to circular motion, stretchable Chinese tongs, hollow tori, nanoscale crossword puzzles... see the applications mentioned in §15) will become routine.

How far can the tinkertoy/tree-like assembly idea of this paper be pushed? What is going to be the ultimate limit? I think the bottleneck is going to be chemical specificity[53]. Suppose you have a reagent which causes some side reactions in addition to the main reaction. For example, in our plan of §13 for planar rigid tinkertoy constructs, we used ozone and pointed out that it also reacts with aromatic "double" bonds in addition to the true double bond we want it to react with, although the rate constant is $\approx 10000\times$ smaller per bond. Also in our multilevel protection based plan of §10 we pointed out various side reactions which could slowly attack our protective polyamide chains. Now for most purposes, most chemists would regard such 10000:1 specificity as pretty good. But, once one's macromolecules have reached the point where they have 10000 vulnerable side sites as compared to only one desired reaction site[54] (remember, with a tree-like assembly, there is always exactly *one* joining reaction site at each stage), that means that the yield for that step has gone down to 50%. Thus, poor chemical specificities will set limits on reachable macromolecule sizes by causing declining, rather than constant, yields[55]. At present, data on reactions of very high specificity is an area that synthetic chemists have largely ignored, but the present paper indicates that it is the key issue. Call it "precision chemistry."

17.1 5-year plan for future development of this area

1. The first task is to continue designing more tinkertoys and experimenting with them in the lab.
   Since

[52]In fact, here is a "proof" that Drexlerian eutactic matter can't be made: Drexler needs to remove surface contaminants adhering to the surface by Van der Waals forces, hydrogen bonds, etc. These sticking energies have order 1 kcal/mole. Meanwhile the bond energies in the eutactic matter are $\approx$ 80 kcal/mole. So how do you get rid of the contaminants? The usual way in ultrahigh vacuum systems is to pump on it for days while baking it at 700 Kelvin. Anyhow, the idea here is that to get eutactic matter we need a temperature high enough that *all* the crud gets unstuck, (say prob $\le 10^{-20}$ of staying stuck) and *none* of the bonds get broken (say prob $\le 10^{-20}$ of breaking a bond). No such temperature exists. End of "proof."

The energy ratio, which I've taken as 80 above, would have to be *at least* 2000 before such a temperature could exist, although actually I think 2000 is a huge underestimate. This provides some insight into why it has been so difficult to achieve ultrahigh vacuums in the lab; the best vacua ever achieved, as far as I can tell, have about 30000 molecules per cc.

[53]Other serious problems, which I suspect will not be *as* serious, may include: solubility (but experience with large dendrimers suggests that may be overcome), declining reaction rates, and purification. More on these subjects later.

[54]Another side of this same issue arises in our scheme of §10 for "stacked protection." One might argue that in a molecule with K protected sites, *all* of which are simultaneously supposed to be deprotected by one level by "operation D," the yields will decline exponentially with K. Of course this is really the same side-site problem in disguise, because failures of the Edman degradation we use for D mainly seem to be caused by occasional random hydrolysis of peptide strands (a side reaction) not by non-occurrence of the degradation itself.

[55]Even with joining reaction yields declining proportionally to $1/N$ for N-size molecules, we still will get total yields $N^{-O(\log N)}$ via balanced tree-like synthetic schemes, i.e. a lot better than an exponential decline, although worse than any power law. ***This is still best possible.***

(a) I am only an amateur chemist;

(b) I have never run any of the reactions proposed here in a lab;

(c) even professional chemists often get surprised by unexpected side reactions and non-working reactions;

it is quite likely that the tinkertoy designs I've proposed in this paper will not work. However, during my design process, I tried as best I could to search the literature to find injurious side reactions, and every time I found one, I always managed without much trouble to modify the design to avoid it[56]. So my feeling is, that the power of modern synthetic chemistry is now so great[57] that, even if some of the details in my designs do not completely work, one will almost certainly have enough flexibility to work around the flaws or create other designs that do work to implement the grand plan.

Good initial experiments to do would include:

(a) just trying to accomplish 2 steps of one of my tinkertoy plans of §12 or §7.2, 11;

(b) trying to make dendrimers (§4.2) of record size to get an idea of the limits one will encounter when trying to make very large treelike molecules.

2. Next, one will gradually have to optimize the yields and reliability of ones's process, and verify[58] the small and medium size structures one produces.

Once one is synthesizing molecules of sufficiently large size and complexity using tinkertoy techniques, purification and proof of structure is quite likely going to become impossible. How can one get rid of side products with almost exactly the correct structure, but some flaw deep inside? You can try to purify at every stage of the synthetic process as best you can, but some percentage of the side products are going to be hard to remove. Certainly optimizing the purification processes – an area we have not discussed at all – is going to be a lot of work; solubility (and the lack thereof!) may be a difficult related issue[59]; and if 20 varieties of tinkertoys are desired with different side groups (much as there are 20 kinds of natural amino acids), then there will be a lot of work to devise the needed protection chemistry for those side groups.

But it seems to me that if good yields are demonstrated on a number of small tinkertoy syntheses, causing these techniques to become well recognized and have well understood reliability behavior, and once the process becomes automated, one would then be justified in extrapolating yields to whatever larger structures makes, and simply rely on the process to work without needing to try to prove structure. In the early days of DNA and polypeptide synthesis, one needed to verify all the structures produced, but nowadays one simply uses the synthesis machine and trusts that it worked.

3. Finally, one will have to build computer controlled machines to perform the synthetic steps needed to produce any shape macromolecules on demand.

The machine will consist of a computer controller, a large number of small jars to hold the subtree chemicals, larger jars to hold reagents, and appropriate mechanisms for combining jars, metering out reagents, and supplying appropriate reaction conditions (pressure, temperature). Specifically, if you want your machine to be capable of building tinkertoy constructs with N tinkertoys, then at the bottom level of the synthesis tree (naively) one would need N small jars. At the next level (assuming a balanced tree synthetic plan) one would need $N/2$ jars (each

[56] Remaining points that may require caution have been mentioned in other footnotes. I have intentionally described more than one alternative method of carrying out almost every chemical operation, to demonstrate the flexibility one has to work around possible problems.

[57] Cheng and Corey 1989 showcase some of the most incredible syntheses that have been accomplished. This sort of work (the synthesis of natural products and the development of new synthetic reactions) will remain entirely unaffected by tinkertoy methods. In terms of chemical sophistication, tinkertoys are far simpler than everything Corey and his community do. (But it is from this simplicity that they derive their power.) In fact I think of these 3 areas: (1) classical synthetic chemistry, (2) "supramolecular chemistry" (the attempt to understand and manipulate intramolecular forces), and (3) tinkertoys/tree-like assembly methods, as disjoint but synergistically related.

[58] e.g. by X-ray crystallography

[59] Large rigid aromatic compounds such as are proposed in §13 are often insoluble. For example copper phthalocyanine (12 fused rings) is soluble in concentrated H_2SO_4 but is insoluble in other common solvents and hexa(peri)benzocoronene ($C_{42}H_{18}$, 13 fused rings) is not known to be soluble in any solvent. However, we the designers are free to adjoin solvent-philic groups at will to our tinkertoys, or to adjoin bulky groups intended to discourage "plate stacking" crystallization and intramolecular bonding forces in general. Thus copper phthalocyanine with two -SO_3Na groups adjoined at antipodal benzene hexagons is "CI direct blue 86," and with four such hydrophilic groups (carbonate groups also work, even if additional aromatic rings are adjoined) we get CI 74220, CI 74320, etc – freely water soluble dyes. Of course water is not the solvent that springs to mind for aromatic compounds, but *this suffices to make my point* that 1 solvent-philic group per 3-4 hexes can solubilize large rigid aromatics. (Other examples from the "Colour index" include CI 68705 [11 fused aromatic hexagons; soluble in xylene], CI 59815 [9 fused hexes, soluble in acetone], CI 59810, CI 69840, etc.) If we instead argue from first principles, note that the gravitational energy corresponding to a 70μm cube of specific gravity 1 falling 1cm is 3×10^{-11} Joules, and its surface energy (assuming solvophilic groups are placed at the nodes of a $10\AA$ square grid drawn on the cube's surface, and each such group contributes a solvation energy of 1 kcal/mole) is also 3×10^{-11} Joules. Hence, provided surface energies work in your favor, solubility is no problem for particles of sizes below 70μm. The trick is going to be to find the proper balance of solvating groups, porosity, and rigidity, while still keeping the chemical properties we need.

one larger than the first level jars since typically one combines the contents of two first level jars to make a second level jar) and at the next level $N/4$, and so on, for a total of $\approx N$ jars. This is a lot of jars, considerably more than in previous synthetic machines. It is possible to reduce the required number by noticing that at the first level of the synthesis, in fact only a few jars are needed, since many of the "different" jars will actually be holding the same chemical. In the event that standardized nanocomponents such as spars, plates, etc., are commonly used, then it would be possible to stock pre-synthesized supplies of them, or alternatively to combine all the "common subexpressions" (identical synthetic subtrees) into common jars, thus again saving jars. The total savings from such tricks might be a factor of ≈ 5. If reaction times are significantly longer than the times required to carry out mechanical operations, then such operations may not need to be parallelized. We've already mentioned the possible usefulness of "solid support" ideas in §7.2.

I believe that this complete 3-step plan, culminating in a working tinkertoy assembly machine, could be accomplished in about 5 years of work. The value of similar mechano-chemical machines has been recognized by the awarding of numerous Nobel prizes to their developers (e.g. Edman, Merrifield, Sanger, Gilbert) and also by their adoption in routine use by laboratories worldwide. It seems to me that the project of automating tinkertoy syntheses could be of similar ultimate value. However, it is more challenging chemically, capable of reaching much larger molecular masses, and more versatile.

18 Abbreviations used in this paper

A, C, T, G : Adenosine, Cytosine, Thymine, Guanine, the 4 kinds of bases that occur in DNA sequences.

ABS : All benzenoid system, see §8.2.

Ar: arbitrary fairly inert aromatic substituent.

DNA: deoxyribonucleic acid.

NP: nondeterministic polynomial time, see Garey & Johnson 1979.

ortho, meta, para : Used to denote relative positions of 2 substituents on benzene rings. Respectively denotes a position 1, 2, or 3 along a benzene ring relative to something; e.g. "para" means antipodal.

O: Asymptotic order notation, often blamed on Don Knuth: $g(n) = O(f(n))$ means a positive real constant c exists so that $g(n) < cf(n)$ when n is sufficiently large. $g(n) = o(f(n))$ means $f(n)$ is not $O(g(n))$; $g(n) = \theta(f(n))$ means both $g(n) = O(f(n))$ and $f(n) = O(g(n))$; $g(n) = \Omega(f(n))$ means $f(n) = O(g(n))$.

Ph: phenyl group or benzene ring. Thus I would denote toluene as Ph-CH_3 and hexachlorobenzene by $PhCl_6$.

PSPACE: the class of problems which can be solved by a Turing machine with polynomially bounded space consumption.

R: arbitrary fairly inert (usually non-aromatic is intended) organic substituent.

X: halide atom {F, Cl, Br, I}.

19 Acknowledgements

James Canary, Nadrian Seeman, Allan Schweitzer, Satish Kumar, P. Kaplan, T.W. Ebbesen, S. Omohundro, and J. Chadi all helped me in roughly decreasing order of importance.

20 References

Adams, Kumar, Martin, Shimamura: Lattice Imaging of Poly(p-phenylene Benzobisoxazole) (PBO) Fiber, Polymer Communications 30 (1989) 285. See also `http://www.gatech.edu/textile/faculty/kumar/kumar.html`.

I.Amato: The apostle of nanotechnology, Science 254 (1991) 1310-1311.

P.L.Anelli & 16 other authors: Molecular meccano I, [2]Rotoxanes and a [2]Catenane made to order, J.Amer.Chem.Soc. 114 (1992) 193-218.

D.B.Annabilino & 16 other authors: Molecular meccano II, self assembly of [n]Catenanes, J.Amer.Chem.Soc. 117 (1995) 1271-1293.

A. P. Alivisatos, K.P. Johnson, X. Peng, T.E. Wilson, C. J. Loweth, M. P. Bruchez, P. G. Schultz: Organization of 'nanocrystal molecules' using DNA, Nature 382 (1996) 609-611.

E.R.Berlekamp, J.H.Conway, R.K.Guy: Winning ways for your mathematical plays, vols 1 & 2, Academic Press 1982.

R.C.Bingham & P.v.R.Schleyer: Chemistry of adamantanes; recent developments in the chemistry of adamantane and related polycyclic hydrocarbons, Springer (Topics in current chemistry # 18) 1971. Also see E.M. Engler and P.v.R. Schleyer: Adamantanes and its Relatives, MTP International Review of Science, Organic Chem. Series One. (Eds. D.H. Hey FRS & W. Parker) Butterworth, London, and University Park Press, Baltimore Vol. 5 (1973) Chapter 7, pp. 239-317.

S.M.Block, B.J.Schnapp, L.S.B. Goldstein: Bead movement by single kinesin molecules studied with optical tweezers, Nature 348 (1990) 348-352. See also K.Svoboda & Block: Cell 77 (1994) 773-784; M.D.Wang, H.Yin, R.Landick, J.Gelles & Block: Biophysical Journal 72,3 (March 1997) 1335-1346.

G.A.Boswell Jr., W.C.Ripka, R.M.Scribner, C.W.Tullock: Fluorination by sulfur tetrafluoride, Organic Reactions 21 (1974) 1-124.

N.Bowden, A.Terfort, J.Carbeck, G.M.Whitesides: Self-assembly of mesoscale objects into ordered two-dimensional arrays, Science 276 (1997) 233-235.

C.M.Buess & D.D.Lawson: The preparation, reactions, and properties of triphenylenes, Chem.Review 60 (1960) 313-330.

F.A. Carey & R.J. Sundberg: Advanced Organic Chemistry (2 vols), Plenum 1990

W. Carruthers: Cycloaddition reactions in organic synthesis, Pergamon (Tetrahedron organic chem. series #8) 1990.

B.Chazelle: Triangulating a simple polygon in linear time, Discr. & Comput. Geom. 6 (1991) 485-524

B.Chazelle & L.Palios: Triangulating a nonconvex polytope, Discr. & Comput. Geom. 5 (1990) 505-536

B.Chazelle, T.Ottman, E.Soisalon-Soininen, D.Wood: The complexity of separation(TM), Proc. 11th Int'l Conf. Automata & Programming (Springer Lecture notes in Computer science 1984) 119-127

B.Chazelle: Computational geometry and convexity, PhD (1980), Yale University (New Haven CT). Also a technical report from Carnegie Mellon University (Pittsburgh PA): CMU-CS-80-150.

B.Chazelle: Convex decompositions of polyhedra, a lower bound and worst case optimal algorithm, SIAM J. Comput. 13 (1984) 488-507.

B. Chazelle & D.Dobkin: Optimal convex decompositions, p. 63-133 in Computational Geometry, (G.Toussaint ed.) Elsevier 1985.

J. Chen & N.C. Seeman: The synthesis from DNA of a molecule with the connectivity of a cube, Nature 350 (1991) 631-633.

X-M. Cheng & E.J. Corey: The Logic of Chemical Synthesis, Wiley 1989.

Eric Clar: Polycyclic hydrocarbons, Academic Press 1964. 2 vols.

Eric Clar: The aromatic sextet, Wiley 1972.

The Colour Index, published by the society of dyers and colourists 1971.

L.R.Croft: Handbook of protein sequence analysis, Wiley 1974.

C.R.Crosby, N.C.Ford Jr., F.E.Karasz, K.H.Longley: Depolarized dynamic light scattering of a rigid macro molecule poly(*p*-phenylene benzbisthiazole, J.Chem.Phys. 75 (1981) 4298-4306.

T.E.Creighton: Proteins, W.H.Freeman 1993.

K.Eric Drexler: Nanosystems: molecular machinery, manufacturing, and computation, Wiley 1992.

K.Eric Drexler: Engines of creation: the coming era of nanotechnology, Anchor Press/Doubleday 1986.

K.Eric Drexler, C.Peterson, G.Pergamit: Unbounding the future: the nanotechnology revolution, William Morrow NY 1991.

T.W.Ebbesen: Carbon Nanotubes, Physics Today (June 1996) 26-32.

M.T.M.El-Wassimy et al.: The S,S acetal to carbonyl transformation using the soft NO^+ or Cl^+ species, J.Chem Soc. Perkin Transactions I,9 (1983) 2201-2203.

E.L.Eliel & S.H.Wilen: Stereochemistry of organic compounds, Wiley 1994.

L.Fieser & M.Fieser: Reagents for organic synthesis, Wiley 1967 (at least 16 additional volumes and a cumulative index meta volume now out).

L.Fieser & K.L.Williamson: Organic experiments, D.C.Heath 1992.

J.F.Friedman, J.Bentley, R.A.Finkel: An algorithm for performing best matches in logarithmic expected time, ACM Transactions on Mathematical Software 3,3, (1977) 209-226.

M.R.Garey & D.S.Johnson: Computers and intractability : a guide to the theory of NP-completeness, W. H. Freeman, 1979.

R.E.Gawley: The Robinson Annulation and related reactions, Synthesis (1976) 777-794.

T.W.Greene & P.G.M.Wuts: Protective groups in organic synthesis (2nd ed.), Wiley 1991.

I.Gutman & S.J.Cyvin: Introduction to the theory of benzenoid hydrocarbons, Springer 1989.

P.J.Hagerman: Flexibility of DNA, Ann. Rev. Biophys. Chem. 17 (1988) 265-288.

H.Hart et al.: Tetrahedron 42 (1986) 1641; J.Am.Chem.Soc. 112 (1990) 3687; J.Org.Chem. 55 (1990) 3412; J.Org.Chem. 56 (1991) 6905, 6912.

Hashida et al.: Bull Chem Soc. Japan 46 (1973) 3263-.

C.Hawker & J.M.J.Frechet: A new convergent approach to monodisperse dendritic macromolecules, J. Chem. Soc. Commun. 15 (1990) 1010-1013.

R.C.Heck: Palladium catalyzed vinylation of organic halides, Organic Reactions 27 (1982) 345-390.

M.K.Herrlein, J.S.Nelson, R.L.Letsinger: A covalent lock for self-assembled oligonucleotide conjugates, J. Amer. Chem. Soc. 117 (1995) 10151-10152.

International tables for crystallography, published by the international union for crystallography, 1988.

N.S. Isaacs: Physical organic chemistry, Longman Sci. & Tech. 1995.

J.D. Joannopoulos, R.D.Meade, J.N.Winn: Photonic crystals: molding the flow of light, Princeton University Press, 1995.

A.W.Johnson: Ylides and Imines of phosphorus, Wiley 1993.

K.E.Johnson & E.L.Green: Rotational energy barrier measured by temperature dependent NMR and modeled using MOPAC, `http://www.pacificu.edu/up/ns/chem/rotation.html` (1997)

M.E.Jung: A review of the Robinson annulation, Tetrahedron 32 (1976) 3-31.

Andy N.C. Kang & D.A.Ault: Some properties of the centroid of a free tree, Information Processing Letters 4,1 (1975) 18-20.

N.Kornblum & D.C. Ifland: The selective replacement of the aromatic primary amine group by hydrogen in aromatic/aliphatic diamines, J. Amer. Chem. Soc. 71 (1949) 2137-2143.

J-M.Lehn (ed. in chief): Comprehensive supramolecular chemistry, Elsevier 1996. 10 volumes plus an index volume. Chapter 25 of volume 10 (pages 777-832) is “Dendrimers” by G.R.Newkome & C.N.Moorefield.

X.J. Li, X.P. Yang, J. Qi, and N.C. Seeman, Antiparallel DNA double crossover molecules as components for nanoconstruction, J. Am. Chem. Soc., 118 (1996) 6131-6140.

D.Lichtenstein: Planar formulas and their uses, SIAM J. Comput. 11 (1982) 329-343.

A.Lingas: The power of nonrectilinear holes, Proc. 9th Int’l. Colloq. Automata, Languages, Programming, Aarhus 1982 (Springer lecture notes in computer science # 140) 369-383.

W.Lipski Jr., E.Lodi, F.Luccio, C.Mugnai, L.Pagli: On two dimensional data organization II, Fundamenta Informaticae 2,3 (1979) 245-260. (Part I was pages 211-226.)

R.K. Mackie, D.M. Smith, R.A. Aitken: Guidebook to organic synthesis, Longman Scientific & Technical 1990.

A.Maercker: The Wittig reaction, Org.Reactions 14 (1965) 270-490.

Jerry March: Advanced Organic Chemistry, Wiley 1992

John E. McMurry: Carbonyl coupling reactions using low valent Titanium, Chem.Rev. 89 (1989) 1513-1524.

H-B.Mekelburger, W.Jaworek, F.Vögtle: Dendrimers, arborols and cascade molecules, Angew. Chem. (Int’l ed. in English) 31,12 (1992) 1571-1576

R.B.Merrifield: Solid phase synthesis (Nobel lecture), Science 232 (1986) 341-347.

C.A.Mirkin, R.L.Letsinger, R.C.Mucic, J.J.Storhoff: A DNA-based method for rationally assembling nanoparticles in macroscopic materials, Nature 382 (1996) 607-609.

J.S.Moore: Hollow organic solids, Nature 374 (1995) 495-496.

R.T.Morrison & R.N.Boyd: Organic Chemistry, Allyn and Bacon 1992

G.R. Newkome, C.N. Moorefield, G.R. Baker: Building Blocks for Dendritic Macromolecules, Aldrichimica Acta 25,2 (1992) 31-37.

G.R. Newkome, C.N. Moorefield, B.D. Woosley, G.R. Baker: Dendrimers, Polymerica Acta. (1996).

Nobel lectures in chemistry 1981-1990 (eds. B.G.Malmstrom, T.Frangsmyr), World Scientific 1992. (Contains lecture by Merrifield & Lehn.)

M.Oki: Recent advances in atropoisomerism, Topics in Stereochemistry 14 (1983) 1-81.

B.A.Patel, C.A.Ziegler et al.: Palladium catalyzed vinylic substitution with carboxylic acid derivative, J.Org.Chem. 42,24 (1977) 3903-3907.

S.D.Razumovskii & G.E.Zoukov: Ozone and its reactions with organic compounds, Elsevier 1984.

Ed Regis: Nano : the emerging science of nanotechnology : remaking the world-molecule by molecule, Little/Brown, 1995.

J.Renegar: On the computational complexity of the first order theory of the reals, parts I,II,III, J.Symbolic Comput. 13,3 (1992) 255-352.

Brian Robinson: The Fischer Indole Synthesis, Wiley 1982. Also, review articles on the same subject by the same author, which are often more convenient because they are shorter, are Chem.Rev. 63 (1963) 373-401 and Chem.Rev. 69 (1969) 227-.

Bruce Robinson & N.C.Seeman: The design of a biochip: a self assembling molecular-scale memory device, Protein Engineering 1,4 (1987) 295-300.

J. O’Rourke: Art gallery theorems and algorithms, Oxford Univ. Press 1987.

J. O'Rourke & K.J.Supowit: Some NP-hard polygon decomposition problems, IEEE Trans. Info. Thy. 29 (1983) 181-190.

D.N.Rubingh & H.Yu: Characteristics of stiff chain macromolecules: poly(n-hexyl isocyanate) in n-hexane, Macromolecules 9 (1976) 681-685.

J.Ruppert & R.Seidel: On the difficulty of triangulating 3-dimensional nonconvex polyhedra, Discrete & Comput. Geom. 7 (1992) 227-255.

W.Saenger: Principles of nucleic acid structure, Springer 1984.

K.H.Saunders & R.L.M.Allen: Aromatic diazo compunds, Edward Arnold (3rd ed.) 1985.

N.C.Seeman & 8 other authors: Synthetic DNA knots and catenanes, New J. Chem. 17 (1993) 739-755.

M.F.Semmelhack: Formation of carbon-carbon bonds via π-allylnickel compounds, Org.React. 19 (1972) 115-198.

J.Shi & D.E.Bergstrom: Assembly of novel DNA cycles with rigid tetrahedral linkers, Angew.Chem. Int'l ed. Engl. 36, 1/2 (1997) 111-113.

V.R.Skvarenko, V.K.Shaleev, E.I.Klabunovskii: Advances in the chemistry of triptycene, Russian Chem. Reviews 43,11 (1974) 951-966.

W.D.Smith: DNA Computers in Vitro and Vivo, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 27, pp. 121-185, 1996. (Unfortunately the two key figures 1 and 2, which are essential for reading the first half of this paper, were deleted by the editors or publisher.)

W.D.Smith: The computational complexity of synthetic chemistry, NECI Tech Report, to appear 1997 (NEC Research Institute 4 Independence Way Princeton NJ 08540 USA).

I.Szele & H.Zollinger: Azo coupling reactions - Structure and Mechanisms, Top.Curr.Chem. 112 (Springer 1983) 1-66

A.L.Thomas: Phthalocyanine research and applications, CRC Press 1990.

D.A.Tomalia, A.M.Naylor, W.A.Goddard III: Starburst dendrimers, Angew. Chemie (Int'l ed. in English) 29 (1990) 138-175.

M.Treacy et al.: A combinatorial method of generating periodic tetrahedral frameworks, NECI Tech.Rept. #96-007 (Jan 16 1996).

M.M.J. Treacy, T.W. Ebbesen, J.M. Gibson: Young's modulus of a single carbon nanotube, NECI Tech Rept 96-034 (1996). Later publication in Nature 381 (1996) 678-680.

V.N.Tsvetkov & I.N.Shtennikova: Structure and rigidity of molecules of aromatic polyamides in solutions, Macromolecules 11,2 (1978) 306-312.

Jiro Tsuji: Organic synthesis with Palladium compounds, Springer Verlag 1980.

H.H.Yang: Aromatic high-strength fibers, Wiley, 1989.

A.F.Wells: Three dimensional nets and polyhedra, Wiley 1977.

A.F.Wells: Further studies of three dimensional nets, Amer. Cryst. Soc. Monograph #8, 1979 (polycrystal book service PO Box 11567, Pittsburgh PA 15238).

O.H.Wheeler & H.N.Battle de Pabon: Synthesis of stilbenes, a comparative study, J.Org.Chem. 30 (1965) 1473-1477.

G.M.Whitesides, J.P.Mathias, C.T.Seto: Molecular self-assembly and nanochemistry, a chemical strategy for the synthesis of nanostructures, Science 254 (1991) 1312-1319.

R.Zhang & W.L. Mattice: Persistence Length of the 'Rodlike' Molecule trans-PBT Revisited Again, Macromolecules 26 (1993) 4384-4385.

Y.Zhang & N.C.Seeman: Construction of a DNA truncated octahedron, J.Amer.Chem.Soc. 116 (1994) 1661-1669.

H.Zollinger: Diazo chemistry (vols I & II), VCH verlagsgesellschaft 1995.

Supplementary figures for footnotes

Supplementary figure 1 (for footnote 44):

```
        H       H    H   H
        C       C    N   C
      // \     / \\ / \ / \\
     C    C---C    C   C    C
...  |   ||  ||   ||  ||    |  ...  Symmetric carbazole linkage.
     C    C   C    C---C    C
      \\ / \ / \ //     \ //
        C   N   C        C
        H   H   H        H
```