The Name Game: A Discussion of Nomenclature

"What's in a name? That which we call a rose by any other name would smell as sweet." …Juliet

Yes, that Juliet. She was complaining (reasonably enough) about the problems that family ties were having on her love life; we know where that led. But in metabolism the names of the plethora of metabolites (those being the small carbon based molecules that are made and consumed in metabolic pathways) can either be a source of massive fear and frustration, or a huge help in deciphering and visualizing the chemistry that is going on. Unlike in organic chemistry, metabolism uses a fairly small number of basic structures in a variety of circumstances, so variations of the same structure crop up in new places. The useful thing is that when the same basic structure is used in a new circumstance, the name will clearly tell you that is the case, if you know what the names mean. This is because unlike the Capulets and Montagues, the names of the metabolites can often (not always) tell you a lot about the structures, and thus make the chemistry and the transitions occurring in a metabolic pathway very obvious instead of shrouded in the mist of a Shakespearian cemetery…

OK, enough with the drama. Specifically, I want to go over two classes of molecules that will massively ease the way to understanding a whole lotta metabolism (as Jerry Lee Lewis would have sung). These two groups are 1) molecules based on glycerol, and 2) dicarboxylic acids. These two small sets of organic molecules have names that clearly indicate their structures, but you’ve got to know how the names tell you this information. I had kind of sort of maybe thought that the prerequisite organic chemistry classes would have covered this ground, especially since they are so popular because premedical majors are required to take them, and that majority block of students would definitely be well served by learning about the molecules that underlie things like energy metabolism and obesity, rather than things like icosahe drane and buckminsterfullerene. But what do I know? Perhaps they did go over this nomenclature, and an overwhelmed undergraduate student simply needs to hear these things twice. After all, repetition is the mother of learning. Whatever your O-chem background, we will go over the name thing here, whether or not you have heard it before. Compared to organic chemistry, the number of new molecules will be small, but learning them will give you a big payoff. Trust me. Remember how I said that you had to learn the one-letter amino acid code? And how some of you might have been sad, mad, or (can’t think of another d word)-d? But every student I have spoken with is glad in retrospect that they did have to learn that code. Well, this is even more directly relevant to the class at hand. The payoff for learning the amino acid code comes mostly after the metabolism class is long over, but the payoff for learning some basic nomenclature will be evident during the class, and especially during your studying for the exams in the class. The one downside to learning the nomenclature is that, unlike the amino acid code, you can not practice with license plates, which is the bestest in the world way to learn the one letter amino acids.

**OK then, what IS in a name?** As you learned in O-chem, the names of molecules can tell us how many carbons are present, what the key functional groups, and where those functional groups are on the molecule. However, there is usually a number of ways to name the same molecule, and so choices and history play a role in the actual name used
in a metabolic pathway. Usually the choices are pretty good ones, but sometimes… not so good. Anyway, a simple example of this idea can be seen with 2-keto-propane, or 2-propanone, which we usually refer to as “acetone” (on the right). The fancier names are completely unambiguous; if you had never seen the pictured molecule, you would still know how to draw it from the name. With the name acetone, you would have to have been told at one time that it is referring to the pictured molecule. So acetone is sort of like a convenient “nickname” for 2-keto-propane. But really, organic chemical nomenclature always uses some sort of nickname conventions so we don’t go crazy. For example, in the more detailed name for acetone, “2-keto-propane”, propane is the standard nickname for the fully saturated three carbon hydrocarbon CH₃CH₂CH₃. The trick is to know the commonly used names for a small subset of molecules. You can trust me that this small amount of learning that will blow open the doors to understanding a lot of metabolism: a little memorization will lead to a big payoff.

Glycerol: when three isn’t a crowd- as you will soon see, or may have seen, the glycolytic pathway has two types of metabolites: those with six carbons and those with three carbons. The six carbon ones are all named in a pretty simple manner, either being derived from glucose or fructose which we talk about in another section. Once the six carbon molecule is broken into two three carbon pieces, then the molecules are named after various three carbon molecules. In metabolism a key prototype structure is glycerol, which is a very convenient nickname for either CH₂OHCHOHCH₂OH, or propane 1,2,3 triol, or probably some other names I don’t know. It is a very simple, symmetrical structure shown in the picture, and many metabolites can be named as different variations of glycerol. Note I don’t say that many metabolites are derived from glycerol, although many people would say this, because in metabolism when we say that we could mean that the metabolite being shown is synthesized in a metabolic pathway that produces it from glycerol, and that is not always the case. Using the glycerol structure as the common nickname (just like we used “propane” above) there are a lot of metabolites that can be named. Sometimes we find glycerol itself within a structure, as in many lipids (tri-glycerides… get it?), but our first encounter with glycerol involves some derivatives. The picture shows the molecule glyceraldehyde. As is often the case, the 1-carbon is the one where the functional group is found. It is a pretty common convention that molecular names should place the functional group for which the molecule is named in the lowest numerical position. This isn’t so important for plain old glyceraldehyde, because you couldn’t tell the difference between “1-glyceraldehyde” and “3-glyceraldehyde” since they would be chemically identical (convince yourself of this if you need to). But what if we have a phosphate group on this molecule? The one shown is “glyceraldehyde-3-phosphate”, a key intermediate of glycolysis. As is the convention, we assign the 1 carbon to the group for which the molecule is named, in this case the aldehyde group at the end. As an exercise, you might try drawing glyceraldehyde-2-phosphate, even though we never encounter this in our pathways…
Glycerol acids are basic… Another common glycerol family includes the carboxylic acids derived from glycerol. Glycerate (the ion form of glyceric acid) is shown in the next picture:

Again, you can see that it is basically glycerol with one of the carbons changed to a different functional group, making it glyceric acid, or glycerate as draw…

**Hyperlink:** WAIT! You don’t know when to use “-ic” or “-ate” in naming acids? Darn those icosahedrane-overfocused O-chem profs! No problem, lets take a brief diversion. There is a very simple rule for designating ionized and unionized organic acids that will serve you for the rest of your illustrious science careers. Ready?

“The -ic acid yields the -ate anion”

What on earth (sorry, Earth) do I mean by that? What it means is that almost universally, protonated organic acids have the “-ic” suffix, and their corresponding anions (when the proton has been given up, that is, when the acid has done its thing) had the “-ate” prefix.

Berry, berry useful. Check out a few examples in the pictures below. Acetic acid is the unionized version of acetate ion. Propionic acid produces propionate when it is ionized. It even works for dicarboxylic acids, but we will get to that later. This is a true rule for any “-ic” acid. For example, fully protonated H₃PO₄, which we will never see in this course, is phosphoric acid, while the often-encountered PO₄³⁻ that we will see, oh, like a billion times, is called phosphate. Similarly, carbonic acid is the fully protonated H₂CO₃, and CO₃²⁻ is called carbonate. (HCO₃⁻ is called bicarbonate, but don’t worry about that). You’d be amazed how long I went teaching this class without simply delineating this basic rule of naming stuff. If biochemistry is a language (it is), then why not teach the basic vocabulary!? Better late than never. Progress is acceptable in the absence of perfection. You will notice that I often choose to use the “ate” form, for the simple reason that in the cell, the pH is held at 7ish, so usually carboxylic (and other) acids are mostly in the ionized form. But you will see that often people use them interchangeably. A notorious example is lactic acid and lactate… but we’ll get to that later. *End of hyperlink*

**Back to glycerol acids**- OK, Let’s do some acid(s).

Way too sixties a reference. Forgive me. Here is glyceric acid, spelled CH₂OHCHOHCOOH. It is technically 2,3 dihydroxypropionic acid, but we call it “glyceric acid”, or glycerate because it is found in the cellular conditions of pH7. To the right is a picture…

So when you hear “glycerate” it is this simple carboxylic acid that the name is based on. Phosphoglycerate is going to be a phosphorylated version of glycerate, that is, glycerate with a –PO₄³⁻ group on one of the carbons. The most common form is 3-phosphoglycerate. Now draw 2-phosphoglcerate as well. This (unlike glyceraldehyde-2-
phosphate) is a metabolite we encounter…somewhere in a glucose-metabolizing pathway…

The next molecule we will consider is called 1,3 bisphosphoglycerate, and it is another critical molecule in glycolysis. You might go “I get it! Now we add one phosphate to the 3 carbon (farthest away from the 1 carbon that names the glycerate, and one more on the… wait a minute! There’s no room for another functional group because the whole 1 carbon IS the functional group!”). Correct. This special case which we will encounter in glycolysis, has one phosphate on the 3 carbon as expected, and the other is actually an acid anhydride of phosphoric acid and the carboxylic acid. Looks like this: Sigh… there are exceptions, but the basic naming based on glycerol is preserved.

**Acetone… another nickname enters the picture**- when I was an organic chemistry student before the continents split off from the Gondwanan land mass, we used acetone all the time to wash our glassware during lab. Mmmmm that stuff smells good, and is cool to the touch due to evaporation! Anyway, the molecule is simple and we referred to it above. To reiterate, acetone is spelled CH₃COCH₃, and can be depicted as shown (there’s a more complete picture above. Both are totally acceptable). Very simple, symmetrical. The simplest possible ketone. The more biological version is dihydroxyacetone, and it is shown next to acetone. You can see that it is also, like glycerol, a three carbon molecule with each carbon bearing one oxygen of one sort or another. The most important acetone-based name we encounter is the dihydroxyacetone phosphate, which is produced in glycolysis when the six-carbon sugar is split into two three-carbon sugars. This name tells you exactly what the structure is. There is a hydroxy group on each of the two available carbons, and one of them is decorated with phosphate. I don’t have a picture of that shown. But you have seen it many times because of its importance in glycolysis, at one corner of the “coat hanger” picture. You will notice that dihydroxyacetone is a completely symmetrical molecule, so we don’t even have to use a number. If you put a phosphate on either –CH2OH the resulting molecule (DHAP) would look and be the same. (Glycerol 1 and 3 phosphate is a bit different: you can distinguish the two forms because the middle carbon is asymmetric). Acetone has the distinction of actually being a (rare) metabolite. As you will learn, it can be generated when the body is making a lot of “ketone bodies” in some metabolic situations.

**A bit of practice**- The following reaction is used in the movement of electrons from the cytosol to the inner regions of the mitochondrion where they can do some good. That detail is not so important, but the names are one of those cases where they sound very different but they are structurally closely related. The nice thing is that now with your bag-of-naming tricks, you can see how the names clearly indicate the structures. You can convince yourselves of this by filling in the structures from the lone names.
Dihydroxyacetone phosphate + NADH + H⁺ ----> NAD⁺ + glycerol-3-phosphate
(use the space above to draw the picture. Actually draw them: convince yourselves!)

You might wonder why the glycerol compound with a single –PO₄³⁻ group is not called glycerol-1-phosphate, to keep the numbers low. This metabolite has the phosphate numbered as 3 because only one of two possible stereoisomers of glycerol phosphate are made in the cell. Still, I hope you can draw these molecules from the names (without worrying about the stereochemistry right now). Try it!

**Ch-co-co! Pyruvate and lactate-** The last molecule produced in glycolysis is our (soon to be) old friend pyruvic acid, or pyruvate, which I like to use because it is anionic at cellular pH. One could also call it propionic acid-2-one, 2-ketopropionic acid, or the acoustic mnemonic for CH₃COCOO- “chicoco”… but we all call it pyruvate. There are not a lot of derivatives of pyruvate, but it is so central to metabolism that it gets its own name. So I include it here. When pyruvate can’t be used up, it is converted into lactic acid, for reasons that will be clear to you when we cover glycolysis. But this section is about names only. That acid is lactate, or lactic acid, and you can see that it is just pyruvate with the ketone group reduced to the corresponding alcohol. But lactate is not called dihyropyruvate, although I suppose one could call it that. It is lactate, but his has a simple and clear relationship to pyruvate.

Take home lesson: Glycerol is on the left, dihydroxy acetone is in the middle, glycerate is on the right! Many (many) metabolites are variants of these simple structures. Know them and know much!

**Dicarboxylic acids, the key to the Krebs cycle-** Ever notice how many metabolites are anionic? Especially in the core energy metabolism reactions, everything is anionic. There are some very cool ideas about why this might be, including speculations that the earliest self-propagating reactions occurred on cationic surfaces that could only hold onto polyanions. Whatever their origin, one of the places that we find abundant anions is the Krebs cycle. It is full of carboxylic acids, and most importantly it is full of dicarboxylic acids. There is a little bit of molecular irony here, because one of the “official names” of the Krebs cycle is the **tricarboxylic acid cycle.** That is because the
very first molecule is indeed a tricarboxylic acid, citrate (or citric acid if you are into protonation), but soon thereafter, it is all about the di’s.

A dicarboxylic acid is just what it sounds like: a molecule with two carboxyl groups, each of which acts as an independent functional group. The simplest one is called oxalic acid, and it is just two carboxyl groups connected together. That’s it! Just like the earlier acids mentioned above, the fully protonated form is called “oxalic acid” and the anionic form is called oxalate. There are a series of dicarboxylic acids, each one carbon longer than the next, to the tune of

\[ \text{O}_2\text{C-(CH}_n\text{)}_{2}\text{-CO}_2^- \quad \text{where } n = 0, 1, 2, 3 \text{ or larger.} \]

For example, our recent example of oxalic acid, \( n = 0 \). Just two carboxyl groups bound back to back. The addition of methylene carbons to the “middle” of the dicarboxylic acids (making \( n \) up to 5 in the examples below) creates a series of dicarboxylic acids, with very “random” (as the young people say now) names. Really, they are historical, but whatever the adjective, they are not systematic and so it would seem hard to extract structural information from the names. The total carbon number, name and structures are as follows

<table>
<thead>
<tr>
<th>Carbon Number</th>
<th>Name</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>oxalic</td>
<td>( \text{O}_2\text{C-CO}_2^- )</td>
</tr>
<tr>
<td>3</td>
<td>malonic</td>
<td>( \text{O}_2\text{C-CH}_2\text{-CO}_2^- )</td>
</tr>
<tr>
<td>4</td>
<td>succinic</td>
<td>( \text{O}_2\text{C-(CH}<em>2\text{)}</em>{2}\text{-CO}_2^- )</td>
</tr>
<tr>
<td>5</td>
<td>glutaric</td>
<td>( \text{O}_2\text{C-(CH}<em>2\text{)}</em>{3}\text{-CO}_2^- )</td>
</tr>
<tr>
<td>6</td>
<td>adipic</td>
<td>( \text{O}_2\text{C-(CH}<em>2\text{)}</em>{4}\text{-CO}_2^- )</td>
</tr>
<tr>
<td>7</td>
<td>pimelic</td>
<td>( \text{O}_2\text{C-(CH}<em>2\text{)}</em>{5}\text{-CO}_2^- )</td>
</tr>
</tbody>
</table>

**Easy as (apple) pie** - This little series only includes the dicarboxylic acids that have carboxyl groups at the ends of linear carbon chains. The amazing thing is that even when only considering this highly restricted collection of structures, a *lot* of metabolism becomes clear when you know how to name this little clan of molecules. The trouble is that these names have no scientific or systematic connection to each other. Perhaps a technical philologist could see a pattern, but for us normal people they are six completely unconnected words. It is the same problem in the Krebs cycle: the poor student encounters a bunch of molecules with totally unrelated names (with the possible exception of succinyl-CoA and succinate, I suppose) that are undergoing discreet biochemical conversions. Usually the student learns the names by complete brute force memorization, but because they don’t know the structural meaning of the names, the Krebs cycle does not become any easier to understand. YUK! That is like making students learn math formulas that do not improve intuition. Hate that! But once you know the structural information that the names are telling you, the Krebs cycle’s chemical logic becomes clear. Alternatively, if you know the chemical logic of the Krebs cycle (you
will, I can promise you that) then assigning the names to the various metabolites is straightforward, once you know how to name them. So, you’d think that the names of the dicarboxylic acids would be taught to pre-meds and other biologically inclined students in O-chem. You’d think. You might also think that an economist would be able to tell us how to get rich. So much for reasonable expectations. It turns out that there is an incredibly simple way to remember the nomenclature for the dicarboxylic acids, and it is a mnemonic. Just like Ch-co-co for pyruvate, which if you haven’t heard it yet, you will.

Ready:

The mnemonic for naming the first 6 dicarboxylic acids is

**Oh Mom! Such Good Apple Pie!**

Pretty quaint, and maybe a bit sexist, because the implication is that moms always cook apple pie, or are at least responsible for registering Pie Awareness. So alternatively we could use a more “street” version:

**Oh Man! Such Good Apple Pie!**

Or a self-obsessed version, also fairly quaint, and one that rhymes:

**Oh My! Such Good Apple Pie**

Or, if you’re a texer-blogger, you might just go for OMSGAP…(right now I am L’ing My F’in A O… not!). For the true pie enthusiast, one can say Great instead of Good. No prob. Greasy for the calorically honest, Glutinous for the cautious, Apricot for Apple-haters, Pastry for Pie-haters, etc. Other variants can be found, I am sure. Make up your own. But the point is, that Oxalic, Malonic, Succinic, Glutaric, Adipic, and Pimelic are the 2,3,4,5,6 dicarboxylic acids, and it is a great (really great) idea to learn these names. But wait, we can even make it easier:

**A mnemonic challenge for young metabolites**- The OMSGAP is a standard mnemonic taught in many organic chemistry classes. Don’t take my word for it… just google “Oh my such good apple pie” and see what you get. So there’s some science history in the use of the full series, and in fact there are three more dicarboxylic acids that are even more obscure, with 8, 9 and 10 carbons, and crazy names (suberic, azelaic, sebacic) that add “sweet as sugar” to the mnemonic. But we never ever encounter those big crazy dicarboxylic acids in metabolic biochem. Never. The ones we REALLY see a lot are just the first 4, Oxalic (2C), Malonic (3C), Succinic (4C), and Glutaric (5C). THOSE are the ones that it really really helps to know. So what we really need is a smaller mnemonic for those for OMSG. We NEED a great mnemonic for OMSG. Please put your young brains to work and come up with one we can deliver to the world. Here are some possibilities

**Otters Make Silly Gestures** (that back-stroke they do)

**Over Me Stands God** (for the pious)

**Oh My! Stan Getz!** (for people who like saxophone)

**Order Me Steamed Garbanzos** (for the busy/hungry)
You can see there is much room for improvement. Please help… send entries to rhampton@ucsd.edu and thanks from current and future generations of metabolites everywhere. Help metabolites young and old and find us a mnemonic.

**Its all Greek to us**- One of the funny things about naming carboxylic acids is that the carbons after the carboxyl carbon have this funny naming system that you just have to live with. For example, the carbon right next to a carbonyl is called the “alpha” carbon, while the carbon two doors down is called the “beta” carbon. The alpha carbon is particularly important because we see α-keto acids in metabolism often, and in anabolism and catabolism of such acids. We also see things like “β-hydroxy-butyrate” and we see β-hydroxy intermediates in metabolism of fats as well. In the case of the dicarboxylic acids, even though there are technically speaking two alpha carbons (the CH2 next door to either carboxyl group) fortunately we don’t see both ever being used this way, so the names that occur instantly give us structural information. For example, let’s think about the Krebs cycle. Even if you haven’t had the Krizzle Cyzzle (Krebs cycle) yet, yo, we can still take advantage of the names therein to flex are new naming muscles. Take for example α-ketoglutarate, often abbreviated αKG. This is not only a main molecule in the Krebs Cycle, but appears in the metabolism of nitrogen and in a variety of other places. Before a student would see (or hear) “alpha keto glutarate” and it would mean almost nothing. It would require rote memorization and have no connection to any other molecule once that memorization was accomplished. That kind of information is highly volatile, at least in this old professor’s brain. But now, with OMSG (Whatever the best mnemonic turns out to be) we can go “hmmm, glutarate (5 carbons) and alpha keto means that there is a carbonyl group (keto) on one of the alpha carbons. So the structure literally falls out of knowing these simple rules. (I used an “alpha” in the typed name because the silly iPad app does not do symbol font. Lame…)

**Backin’ your play with skills**- But perhaps more importantly, ANY molecule with “glutarate” in it anywhere is subject to your incisive analysis. Par ejemplo, one of the key molecules in sterol metabolism, as well as ketone body formation, is HMG-CoA. HMG stands for 3-hydroxy-3-methyl-glutarate, and HMG-CoA is the version of this carboxylic acid that has CoA-SH attached to it. Before knowing the names: “huhhhhhgggg?” After knowing the names: “hmmm… well glutaryl-CoA will be glutarate connected to CoA, like we have seen for acetyl-CoA, and the actual acid (HMG) will be 3-OH-3-methyl-glutarate. So, five carbon dicarboxylic acid with both an OH and a methyl group on the three carbon. See? Now you have sweet naming skills!
Squaring the circle- When we study the Krebs cycle, you will see that the cycle starts by generating a six carbon tricarboxylic acid called citric acid, that is then subject to simple chemical steps that cause the loss of carbons as CO₂, and the reduction of carriers to generate electrons for use in energy production. As I said above, the names of each molecule are all seem unrelated, but in fact a number of them are based on the dicarboxylic acids, and so you can tell not only what they look like but where they arise in the Krebs cycle because the number of carbons will be clear and unambiguous. This isn’t the case for all the Krebs intermediates, but for enough of them that the cycle is a lot easier to get your head around when you know this. Here is the list in order of the cycle: citrate, isocitrate, α-ketoglutarate, succinyl-CoA, succinate, fumarate, malate, oxaloacetate

So here we go: the old way would be “eight completely uninformative and complex names I need to memorize”. The new way is citrate, isocitrate: six carbon starting molecule and an isomer α-ketoglutarate: Wait! I know this, it is a dicarboxylic acid with 5 carbons, and a keto group at the alpha (one away from a carboxyl) position! succinyl CoA: I know this one too! Based on a dicarboxylic acid with four carbons, and one of the carboxyl groups is esterified to CoA-SH just like in other CoA adducts! Succinate: Lost it’s CoA, but still a dicarboxylic acid with four carbon (S)! Fumarate, Malate: four carbon dicarboxylic acids we haven’t named because they are fancier, one with a double bond, one with an alpha OH. So no luck there… Oxaloacetate (OAA): Wait! Oxalate is a dicarboxylic acid with two carbons (the O in Oh mom…), and acetate is acetic acid, so it is a structure composed of these two parts, a group derived from oxalic acid and an acetate (see pic….)

Not only do the names thus store a lot of structural information, but they help keep the order of the Krebs cycle less mysterious: You go from six carbons, to five, to four, and then back to six to start all over again. In fact, the names alone will tell you where the CO₂s are being released. Why? Because each time a CO₂ is released, the carbon number of the remaining molecule drops by one. Since we start with the six carbon isocitrate, the transition to 5 carbon a-ketoGLUTARATE (omsG… like that) must mean a CO₂ is lost.
Similarly, the transition from 5 carbon αKG to SUCCINYL-CoA (omSg…) must mean that another CO$_2$ is lost. So you know over half the Krebs cycle, IN THE CORRECT ORDER! Cool! Now go forth, and name some NAMES!