Hidden Markov models

You can think of them as a causal graph. Think of the directed graph of the Markov chain.. but then imagine that this is atop view, and you're actually looking at an enormous tower. The members of each layer are mutually exclusive, so when you know one you know all of them.

So, if you want to know the probablility of a particular state in a particular layer, and you know which of the states in the previous layer happened, then you know whether or not every proposition that connects direcly to the truth of the statement of interest, and th ewhole rest of the tower above it is D-separated. So beleiving that the previous state is the only possible cause of the current state, or that t acts the same qway on its own as in conjunction with anything else, is the same as believing you have a Markov chain. So of course I do not believe that protein ssequences are actually Markov chains.

And there are emitted symbols. Each symbol emission can only be caused by states in a particular layer, so of course knowing the truth or falsity of every statement in the layer D-separates the emission from the rest of the tower, including the previous symbol emissions.

Sequence alignment using gap penalties the way ClustalW does is a hidden Markov model. The markovian model of evolution is for a single position; it's not a walking-down -the-chain Markov model. For pairwise alignment, practically speaking, it's just probabilities of pairs. (multi0le alignment, I don't know - what's the probability of a column? No wonder those programs build trees). So it's a three-state HMM - emit a pair, emit a gap iun the top sequence,or emit a gap in the bottom sequence. And that has the same property of having gap opening penalties (tranisition to one of the gap states from the pair state) vs different gap *extension* penalties (transition from one of the gap states to itself). They'd be gap bonuses if they were more likely than 50%.

Okay, I still have unanswered questioons - when ClustalΩ uses an HMM, *what is the content of the HMM*?

One thing it might be is an extension of the concept of a profile, to include gaps. When you transition from match state to match state, you actually end up in a different match state, with the amino acid frequencies of a column in the profile. (probabilities of pairs after all are a function of both background probabilities and transition proibabilities - but I"m not sure that's exactly what happens in a profile).

I have heard the phrase "poor quality alignments" - what if I'm building a profile from bad alignments? I don't think they'll be *that* bad--after all, I got decent alignments with Gonnet--but what if they're kind of bad? Are they still appropriate input, or are HMM's terrible unless you build them from hand-made perfect alignments? Do they amplify the errors? It's a complicated question - it's implicit, it's not something you can just check in the math! Unless you invent a mathematical object called the "error", I guess that's how you'd do it.

But now Krogh and Hughey are saying that you dont actually *need* an alignment. I have no idea how you'd produce it without it--or *with* it--what about the HMM's in HHOMP? Were they produced from the alignment?

PSIPRED does use a circular topology HMM. Profile HMM's for sequences of fixed length, circular HMM's for repeating patterns of variable length.

Questions.  
Can I just download HMM's from HHOMP and use those?  
- yes

Are the HMM's from HHOMP the profile HMM's, or are they something different?  
- this is made by the HHALIGN guys and they have the same file extensionas hhalign uses, .hhm.

What kind of HMM's does ClustalΩ use? Are they the same kind as HHOMP? If so, why wouldn't HHOMP align using their HMM's, why would they use Kalign instead?  
- they probably are since ClustalΩ uses HHalign which was made by the HHOMP people. But ClustalΩ came out two years after HHOMP so they had to use Kalign

Do the HMM's HHOMP uses need an alignment to be trained? What about the HMM's I can produce with ClustalΩ? Why do the HHOMP people align their sequences at all?  
-- I have no idea. The language in the papers suggests that HHalign derives its HMM's from multiple sequence alignments - but Krogh says you can make profile HMM's from unaligned sequences! So why do they do it?

Would HMM alignment give better alignments?  
-- Probably, since it can get both the loops and the strands right, it'll act different in different parts. It worries me that there's a Gonnet matrix in the code... what are these "pseudocounts" that it is used for? But it definitely seems worth a try.