No Atoms

So far
• atoms → coarse grained → lattices

Today – the holistic lecture
• from reaction kinetics to substitution matrices

What if we forget atoms and residues?
• Kinetics / dynamic systems
  • A → B breakdown of A, \( \frac{d[A]}{dt} = k[A] \)
  • foxes and hares \( \frac{dn_h}{dt} = \alpha n_h - \beta n_h n_f \) and \( \frac{dn_f}{dt} = \gamma n_h n_f - \delta n_f \)
    \( n_h, n_f \) number of hares and foxes
  • complicated kinetics – bacterium eats 10 different nutrients, makes 10 waste products, interconversion of nutrients
Plan

- simplest systems
  - one or two reactants
- treatment of more complicated systems
- transition matrices in sequences

- different approach
- handling very low probabilities
Simplest systems

- one species breakdown / radioactive decay
- $A \rightarrow B$ or $A$ disappears
- philosophy
- we know the average disappearance of $A$
- Each molecule has an equal chance of breaking down

$$\frac{dA}{dt} = -kA$$

$$\frac{dt}{dA} = -\frac{1}{kA}$$

$$t = -\frac{1}{k} \ln A - \frac{1}{k} \ln c$$

$$\ln A - \ln c = -kt = \ln \left( \frac{A}{c} \right)$$

$$\frac{A}{c} = e^{-kt} \quad \text{so } A = ce^{-kt} \quad \text{or } A = A_0 e^{-kt} \quad \text{not unexpected}$$
forward and backward reactions

\[ k_1 \]

\[ 2A \rightleftharpoons B \] so \[ 2A \rightarrow B \] and rate of disappearance is \( k_1 A^2 \), rate of appearance is \( k_2 B \)

\[ k_2 \]

\[
\frac{dA}{dt} = -2k_1A^2 + 2k_2B \quad \text{and} \quad \frac{dB}{dt} = k_1A^2 - k_2B
\]

Theme

lots of processes are easiest to describe in differential form (rate of change)
These are easy enough to do by hand
Make it more complicated
An enzymatic reaction

\[ E + S \overset{k_1}{\underset{k_2}{\rightleftharpoons}} ES \rightarrow E + P \]

\[
\frac{dE}{dt} = -k_1 E \cdot S + k_2 ES + k_3 ES
\]

\[
\frac{dS}{dt} = -k_1 E \cdot S + k_2 ES
\]

\[
\frac{dES}{dt} = k_1 E \cdot S + k_2 ES - k_3 ES
\]

\[
\frac{dP}{dt} = k_3 ES
\]

let us rewrite..
\[
\begin{align*}
\frac{dE}{dt} &= -k_1 E \cdot S + k_2 ES + k_3 ES \\
\frac{dS}{dt} &= -k_1 E \cdot S + k_2 ES \\
\frac{dES}{dt} &= k_1 E \cdot S + k_2 ES - k_3 ES \\
\frac{dP}{dt} &= k_3 ES 
\end{align*}
\]

- We have a matrix form
- What is \( k_1 E \cdot S \)? (and next terms)
- you would usually say velocity vector \( \mathbf{v} \)
- we can describe everything as \( \mathbf{s} = \mathbf{Nv} \)
General approach to kinetics

• "differential form" of kinetics
• applicable to most reactions

How is it helpful?

• \( \frac{dA}{dt} \) is a velocity in one dimension
• velocity of A depends on where A is, B is, ...
• how to predict behaviour of system?

• For some initial \( A_t \) say \( A_{t+\Delta t} = A_t + v\Delta t = A_t + \frac{dA}{dt}\Delta t \)
  • numerical integration exactly as in Newtonian dynamics
  • do the same for \( A, B, C \) ...
• Not just in this lecture – maple, matlab, deSolve in R, ..
Even more general

- We have a number of states $i, j, \ldots$ starting materials, products, intermediates
- we have a finite amount of material
  - use the term probability $p_i$ for convenience and consistency
- $p_i(t + \delta t)$ depends on initial value, flux in and flux out

\[
p_i(t + \delta t) = p_i(t) + \delta t \sum_{i \neq j} k_{ji} p_j(t) - \delta t \sum_{i \neq j} k_{ij} p_i(t)
\]

$k_{ab}$ is rate constant for $a \rightarrow b$

- or given a set of reactants and a matrix of $k$'s (rate matrix)
  - we can model the system
- if we say $v_{ij} = p_i k_{ij}$ what is the meaning of equilibrium? Every $v_{ij} = v_{ji}$
  - for an arbitrarily complicated system
    - I can find the set of $p \ldots$ equilibrium concentrations
the master equation

- in chemical modelling, physical processes, work with master equation

- modelling in engineering
  - put all components and possible routes into numerical bucket
  - find steps which are bottle-necks
  - effect of alternative pathways, think of multitude of protein folding pathways

Last property
- the state at $t + \delta t$ depends on state at $t$ and rate constants
- no dependence on previous states = Markov process

- what is the connection to sequences and mutations?
Markov processes and mutations

- First – more general idea of transition matrices / Markov Chains
- My system is described by a vector of probabilities – think amino acids at a site

\[
p = \begin{bmatrix}
p_A \\
p_G \\
p_C \\
\vdots
\end{bmatrix}
\]

for ala, gly, cys, ...

- \( p_{AB} \) probability of a transition AB but we have lots of them
A Markov transition matrix

\[
\begin{bmatrix}
D & E & \ldots & W \\
D & p_{DD} & p_{DE} & \ldots & p_{DW} \\
E & p_{ED} & p_{EE} & \ldots & p_{EW} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
W & p_{WD} & p_{WE} & \ldots & p_{WW}
\end{bmatrix}
\]

Only valid for short times
- D→E  OK
- D→S→T→A→D→E  something different

In Markov / probability framework rows sum to 1
Applying a matrix

• imagine three kinds of amino acid, $P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$

• population $E, D, W = 0.4, 0.4, 0.2$

• at time $t + \delta t$

\[
\begin{bmatrix}
0.7 & 0.2 & 0.1 \\
0.3 & 0.6 & 0.1 \\
0.1 & 0.1 & 0.8
\end{bmatrix}
\begin{bmatrix}
0.4 \\
0.4 \\
0.2
\end{bmatrix}
= \begin{bmatrix}
0.7 \cdot 0.4 + 0.2 \cdot 0.4 + 0.1 \cdot 0.2 \\
0.3 \cdot 0.4 + 0.6 \cdot 0.4 + 0.1 \cdot 0.2 \\
0.1 \cdot 0.4 + 0.1 \cdot 0.4 + 0.8 \cdot 0.2
\end{bmatrix}
\]

• gives us the new state of the system

• is this a substitution matrix?
comparison with a substitution matrix

blosum62:

|   | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
| A | 4 | -1| -2| -2| 0 | -1| -1| 0 | -2| -1| -1| -1| -2| -1| 1 | 0 | -3| -2| 0 |
| R | -1| 5 | 0 | -2| -3| 1 | 0 | -2| 0 | -3| -2| 2 | -1| -3| -2| -1| -1| -3| -2| -3 |
| N | -2| 0 | 6 | -1| 3 | 0 | 0 | 0 | 1 | -3| -3| 0 | -2| -3| -2| 1 | 0 | -4| -2| -3 |
| D | -2| -2| 1 | 6 | -3| 0 | 2 | -1| -1| -3| -4| -1| -3| -3| -1| 0 | -1| -4| -3| -3 |
| C | 0 | -3| -3| -3| 9 | -3| -4| -3| -3| -1| -1| -3| -1| -2| -3| -1| -1| -2| -2| -1 |
| Q | -1| 1 | 0 | 0 | -3| 5 | 2 | -2| 0 | -3| -2| 1 | 0 | -3| -1| 0 | -1| -2| -1| -2 |
| E | -1| 0 | 0 | 2 | -4| 2 | 5 | -2| 0 | -3| -3| 1 | -2| -3| -1| 0 | -1| -3| -2| -2 |
| G | 0 | -2| 0 | -1| -3| -2| -2| 6 | -2| -4| -2| -3| -2| 0 | -2| -2| -3| -3 | 0 |
| H | -2| 0 | 1 | -1| -3| 0 | 0 | -2| 0 | -3| -3| -1| -2| -1| -2| -1| -2| -2| 2 | -3 |
| I | -1| -3| -3| -3| -1| -3| -3| -4| -3| 4 | 2 | -3| 1 | 0 | -3| -2| -1| -3| -1| 3 |
| L | -1| -2| -3| -4| -1| -2| -3| -4| -3| 2 | 4 | -2| 2 | 0 | -3| -2| -1| -2| -1| 1 |
| K | -1| 2 | 0 | -1| -3| 1 | 1 | -2| -1| -3| -2| 5 | -1| -3| -1| 0 | -1| -3| -2| -2 |
| M | -1| -1| -2| -3| -1| 0 | -2| -3| -2 | 1 | 2 | -1| 5 | 0 | -2| -1| -1| -1| -1| 1 |
| F | -2| -3| -3| -3| -2| -3| -3| -3| -1 | 0 | 0 | -3| 0 | 6 | -4| -2| -2 | 1 | 3 | -1 |
| P | -1| -2| -2| -1| -3| -1| -1| -2| -2| -3| -3| -1| -2| -4 | 7 | -1| -1| -4| -3| -2 |
| S | 1 | -1| 1 | 0 | -1| 0 | 0 | -1| -2| -2| 0 | -1| -2| -1 | 4 | 1 | -3| -2| -2 |
| T | 0 | -1| 0 | -1| -1| -1| -1| -2| -2| -1| -1| -1| -2| -1 | 1 | 5 | -2| -2| 0 |
| W | -3| -3| -4| -4| -2| -2| -3| -2| -2| -3| -2| -3| -1 | 1 | 4 | -3| -2| -2| 1 | 1 |
| Y | -2| -2| -2| -3| -2| -1| -2| -3| 2 | -1| -1| -2| -1 | 3 | -3| -2| -2| 2 | 7 | -1 |
| V | 0 | -3| -3| -3| -1| -2| -2| -3| -3| 3 | 1 | -2| 1 | -1| -2| -2 | 0 | -3| -1| 4 |

Andrew Torda

24/10/2018 [13]
where do blosum and PAM come from?

- take related sequences – no alignment errors
- count mutations (transitions) for each AB pair
- correct for the amount of A, B ($p_A, p_B, \ldots$)

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Sequence 3</th>
<th>Sequence 4</th>
<th>Sequence 5</th>
<th>Sequence 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWYIR CASILRKIYIYGPV GVSRLRTAYGGRK</td>
<td>NRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WFWYR CASILHLYIRSPA GVGSITKIYGGRK</td>
<td>RNG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYVR AAAVARHIYLRKTV GVGRRLKVHGSTK</td>
<td>NRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYFIR AASICRHLYIRSPA GIGSFEKIYGGRR</td>
<td>RRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYYR AASIARKIYLRQGI GVGGFQKIYGGQR</td>
<td>RNG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WFWYR AASVARHIYMKQV GVKLKNLYGGAK</td>
<td>SRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYFIR AASVARHIYMKQV GVKLKNLYGGAK</td>
<td>SRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYYR TASIARRLYVRSPG GDALRLVYYGSK</td>
<td>RBG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYYV TASIARRLYIRSPG GVGRRLVYGGNK</td>
<td>RRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYYR TASIARRLYIRSPT GVGSMTKIYGGRR</td>
<td>RNG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYYR AASTARHLYLRGGA GVGSMTXIYGGQR</td>
<td>RNG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WYYYR AAALLRKYIOGFV GVNLRTHYGGK</td>
<td>DRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
transition matrix versus blosum (PAM, JTT, Gonnet, ..)

• philosophically related – not the same

• a substitution matrix is a log-odds creation - $\log \frac{n_{AB}^{obs}}{n_{AB}^{exp}}$
  • scaling does not matter

• a transition matrix is based on formal probabilities
  • if we have a composition vector $\mathbf{v}$ elements sum to 1
  • after multiplication, still sum to 1

• similarities ...

• application to longer times
longer times

- transition matrix tells me about some change $\Delta t$
  - $p_{t+\delta t} = Pp_t$ for composition vector $p$ and matrix $P$
  - then at next time
    - $p_{t+2\delta t} = Pp_{t+\delta t}$ or $PPPp_{t+\delta t}$

- to go to longer times, repeatedly multiply the matrix
- what happens? diagonal elements represent conservation ($p_{AA}$)
  - probability mass moves away from diagonal

- basis of PAM 100, PAM 200 ... substitution matrices

- when doing alignments, one should use the correct substitution matrix
infinite time

- I have a system described by probability of states $p$
- I repeatedly multiply by a realistic $P \ldots P^\infty p$

- does my distribution disappear? become flat?

- with infinite time everything becomes equally likely

- realistic? No
  - alignments become less reliable with evolutionary time
Summary so far

- chemical kinetics, mutation trajectories, fox + hare populations
  - examples of dynamic systems – very similar methods to treat them
  - allows one to treat complicated kinetics
  - usually simulated by numerical integration
- systems biology problems? the same?
  - sometimes yes – sometimes neglect conservation of mass and formal treatment

- a Markov process state at $t + \delta t$ depends on state $t$
  - do not talk about second order or $n^{th}$ order processes

- everything so far depends on bulk properties
  - what happens if you only have a few molecules? small numbers? Last lecture