Lattice Models

Andrew Torda, June 2008, 67.912 Struktur & Simulation

So far

- classify models by detail

detail	type	properties
high	quantum mechanical	very physical
	atomistic	some approximations, mostly physical terms
low	coarse grain	crude functions, approximations, often non-physical terms

Another important property

- continuous vs discrete

Continuous versus discrete

Continuous models
- coordinates (and other properties) take on any value
- typical properties
  - can take derivative with respect to coordinates
  - energy defined almost everywhere

Discrete
- coordinates (maybe more) are limited to certain values
  - think real/float versus integers
- examples
  - weather forecasts, oceanography, wind tunnels
  - finite element methods (engineering)
  - statistical mechanics (Ising model)

Discrete proteins

How do we make proteins discrete ?

- most common
 - lattices, grids, (Gitters)

Who likes lattices ?

- crystallographers invent and name them
- calculators

Proteins on lattices

- different to wind tunnels (usually)
  - attempt to model details of real world
- proteins
  - sometimes picture of real world
  - more common – a very simple model to analyse some property

Putting a protein on a lattice

- Put down a protein and find nearest grid point
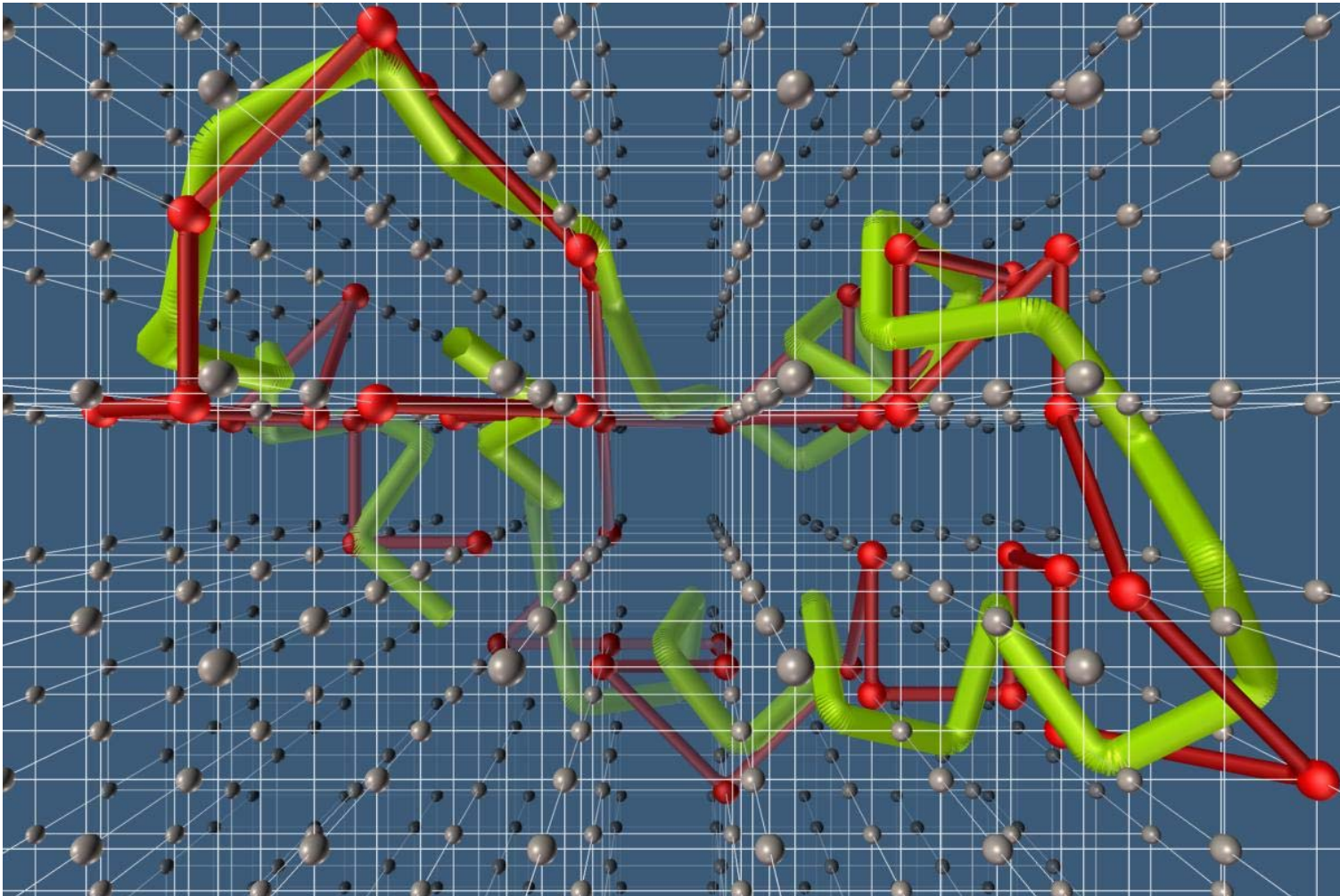
Lattices, errors

What would the error be on a lattice ?

- for 1 Å, should be ½ Å

- can be made arbitrarily small

- what if two continuous residues map to one point ?

  - worse error

- not the only (or best) criterion

  - first, what would our energy look like ?
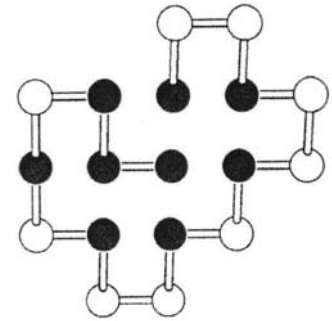
Energy functions

Two philosophies

1. mimic approximation to real energies
   - earlier picture

2. simpler approach
   - use continuous space for realistic simulations proteins

   - use simple model for some topic of interest

Typical energy function

$$U = \sum_{i<j} c_{ij} \Delta(\vec{r}_i, \vec{r}_j)$$

- c_{ij} is some parameterisation constant for types i and j
- switching function

$$\Delta(\vec{r}_i, \vec{r}_j) = \begin{cases} 1 & \text{if } i - j \neq 1 \text{ and } |\vec{r}_i - \vec{r}_j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

Why simple energy functions ?

Simple functions (contact terms)

- some residues like to interact with each other
- will be happiest when the most favourable contacts are made (like a real protein)
- can reproduce very specific structures
  - interactions can be anything you want
- gross properties like hydrophobic packing

Reduced alphabets

Typical question – we want to guess
- how does folding time depend on size ?
- how much hydrophobic area is exposed for some sequence

Do we need 20 amino acids ?
- for general principle, consider 5 or 6 residue types
  - charged - (asp, glu)
  - charged + (lys, arg)
  - polar (thr, ser, gln, asn)
  - hydrophobic aromatic (tyr, phe, his, trp)
  - hydrophobic aliphatic (ala, leu, val, ile, met, cys)
  - special (gly, pro)
- simpler
  - two residue types (hydrophobic / polar, HP)
- Say that protein structure is dominated by hydrophobic collapse
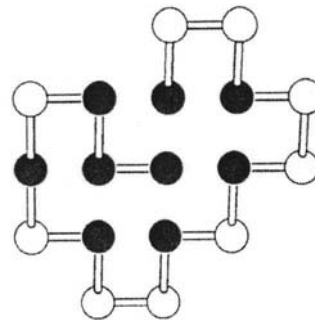 - two residues are really enough for many calculations

Reduced dimensions

Do I care about specific real proteins ?
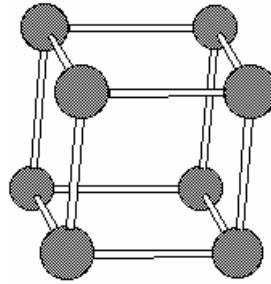
- not always

Is there a simple system which looks like a protein ?

- two dimensional protein

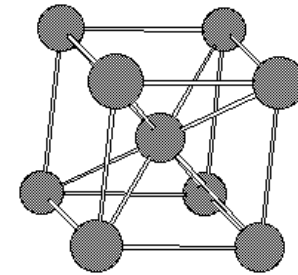- very very simple protein ?

- 2-D, HP model
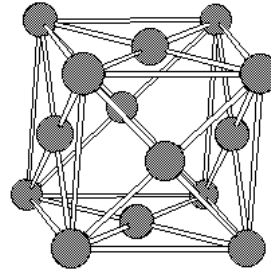
Different types of lattice
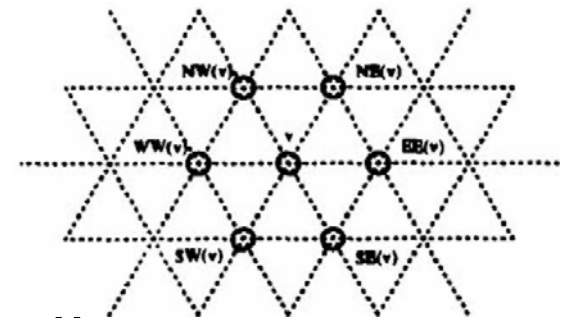
- simple cubic lattice

- body centred cubic

- face centred cubic

- triangular 2D / 3D

most important difference ?

- score functions count contacts
- how many neighbours does each have ?

Why are lattice calculations so fast ?

- normal code
    for each particle
        for each other particle
            is it a neighbour ? calculate energy
- lattice code
    for each particle
        set up list of neighbour cells (often 6, 8, ..)
        look if neighbour is occupied
- what if we have a very realistic system ?
    - all distances can be precalculated
        - 1 unit is 3.8 Å or 0.5 Å or …
    - no more square roots $(x)^{½}$, cutoffs, …

Calculations

We have some machinery, what kinds of calculations ?

- simulation (brief now, more later)
- others

Simulating on a lattice

- we do not have gradients of our energy terms (not much help if we do)
- we do know the energy of a configuration
- calls for Monte Carlo

Lattice simulations

Monte Carlo

- apply normal conventions
- take a step
- calculate energy
  - accept / reject according to Metropolis criterion
- what would our moves look like ?
  - anything reasonable
  - from one starting point, should (eventually) be able to reach any other
  - typical moves ..

Move sets

end move

corner flip

crankshaft

reptation
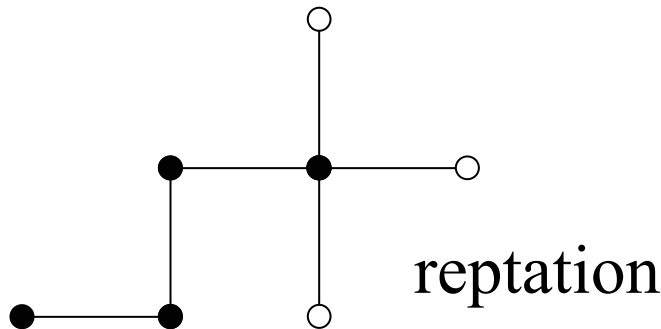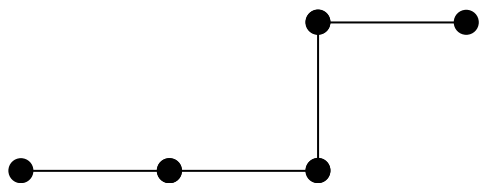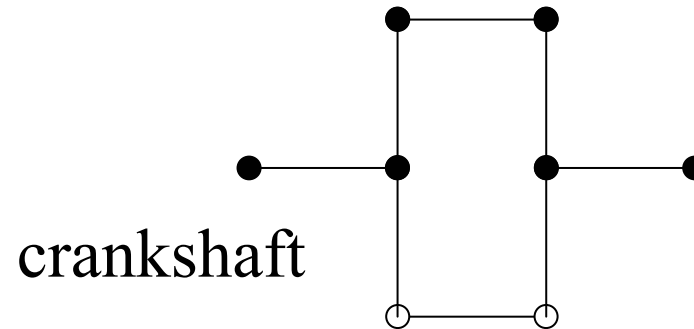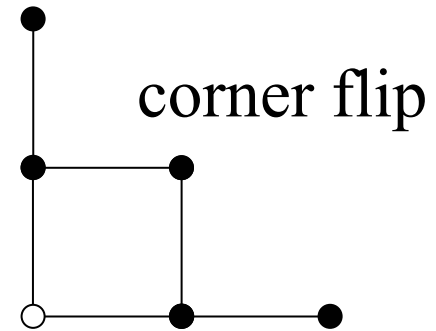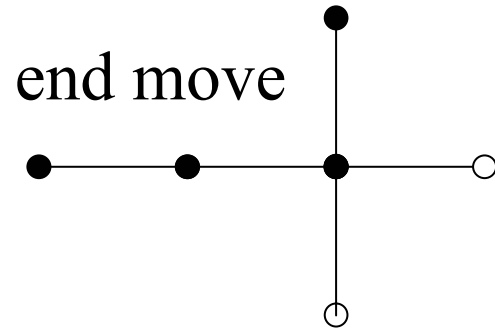
What can we get from simulating

- Take a system usually < 100 residues
- start from
  - random configurations
  - extended configurations
  - misfolded configurations
- run for 10^6 or as many steps as you can afford
  - does a simulation always find a similar minimum energy ?
  - what is the energy spread of misfolded structures ?
  - are there many similar low energy structures ?
  - are there a large number of different low energy structures ?
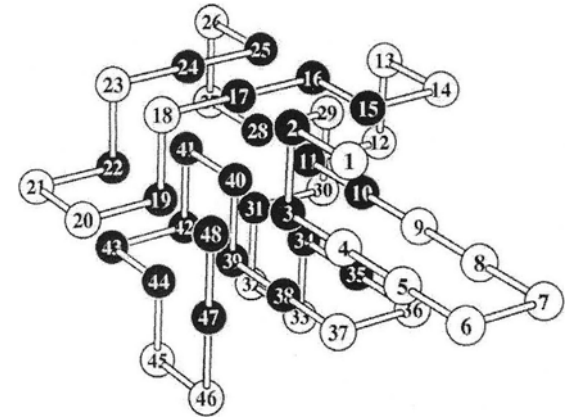
Results from simulations

From a 3D HP model, typical structure
- features ?
  - hydrophobic residues in middle

Compare with MD simulation
- biggest simulation in literature
  - 36 residues
  - months of cpu time
  - does not find global minimum

More on simulations later…

diagram from Dill, … Chan, Protein Sci 4, 561-602 (1995), Principles of Protein Folding …

Unique possibilities

Big problem with atomistic systems
- for any system more than about handful of residues
  - nearly impossible to visit all conformations
- for more than about 10 residues (maybe 15 or 20)
  - little evidence that the global minimum can be found

Lattices
- exhaustive enumeration (visit all possibilities)
  - configuration
  - sequence
- location of optimal structure

Exhaustive enumeration of conformations

Why bother ?

- define almost all the stat mech properties of a system
- remember partition function
- summation over all conformations

$$Z = \sum_i e^{\frac{-E_i}{kT}}$$

We can find things like

- free energies

- distribution of energies

How many configurations are there ? 2D HP model

- 16 residues in 2D is no problem
- in 3D, about 3 x 3 x 3 feasible

length	num configurations
14	110 188
16	802 075

Exhaustive enumeration of sequences

20 amino acids

- too hard

5 or 6 amino acids

- quite realistic, but difficult

HP model ?

- 16 residues is easy (65 536 sequences)
- with this machinery, what can we do ?

Example question

- folding
  - what are driving forces ? (hydrophobic collapse, HP)
  - what is first to form (local or long range ?)
  - how smooth is the folding pathway
  - more later
- evolution
  - more later
- do all protein sequences fold ?
- sequence vs structure space ?

Do all proteins fold ?

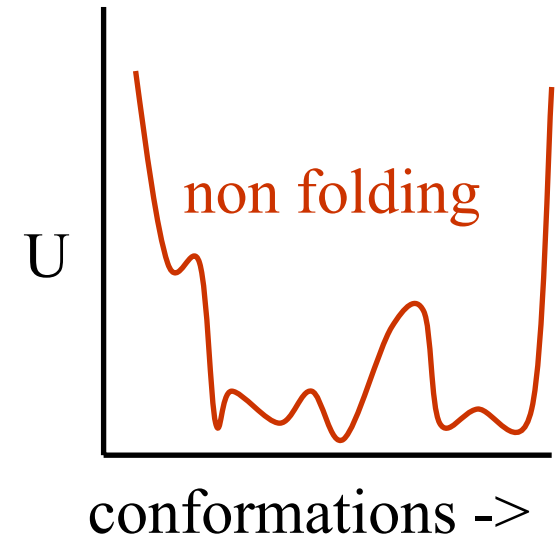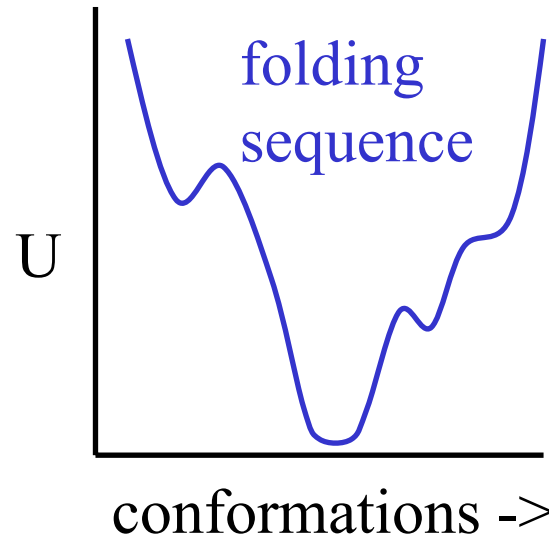If I take a random amino acid sequence, is it a protein ?

- experiment ? less than 1 in 100 fold

- MD simulation

  - cannot even fold one protein

Lattice models

- well studied problem

Definition

- important property

  - folding vs non-folding

U   folding sequence

conformations ->

U   non folding

conformations ->

Folding versus non-folding

Non-folding in a lattice model

- find a sequence
    - visit all conformations
    - rank energies
- how many different conformations have the lowest energy ?
- how many have energy within kT (could be visited at T) ?

Answers ?

- most random sequences do not fold
- intuitive example
    - a very very hydrophobic sequence is happy as long as it is compact
    - there are many ways to make it compact
- agrees with experiment

Sequence versus structure space

From earlier lectures

- different sequences may fold to nearly same structure
    - large number of different sequences known for
        - globins, β-sandwiches, …
- different structures ? usually have unrelated sequences

Can we see this from MD ? No.

From lattice

- for each configuration
    - try every sequence and see if it is an energy minimum
- see how many sequences like each structure

Favourite structures

Some structures are the minimum energy for many sequences

- in a 3×3×3 HP model, there are 100's of sequences which like this structure

- some structures are popular, some much less so

- in principle, totally agrees with nature

  - exact numbers have no meaning

from Li, Tang & Wingreen, Proteins, 49, 403-412 (2002), "Designability of protein structures"
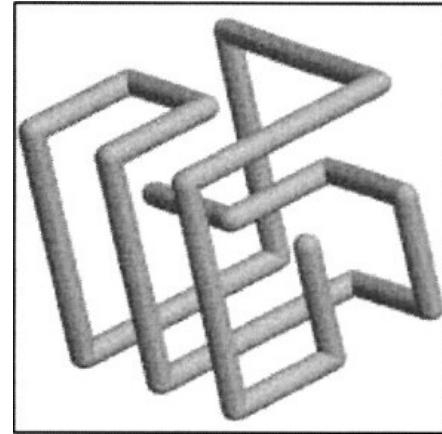
Problems and limitation of lattices

- statistical mechanics are completely valid, but…
- loss of detail
  - resolution is obvious
  - interpreting in physical (or structural) requires faith
    - example of α-helix in 2D
  - whole structural properties may be lost
    - chirality ? chirality of a helix
- discretisation
  - energies and configurations are discrete
  - if a property depends on number of states, results will be model-dependent
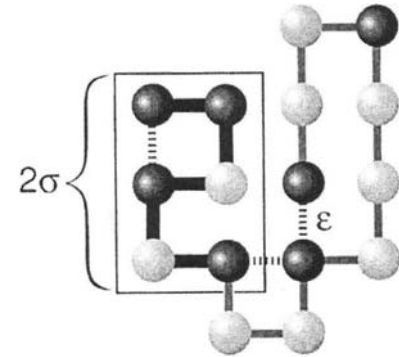
Relating lattices to the real world

- simple models and reduced alphabets
  - only trends are believable
  - some trends can be tested
    - how do results change with 2 versus 3 amino acids ?
- for detailed models,
  - dependence on lattice type and resolution

Artefacts

Susceptibility to artefacts ? Examples
- dependence on alphabet size
    - how popular is a structure may depend on alphabet size
    - in simple alphabets and energies, there are less foldable structure
        - more complicated models make lowest energy more unique
- properties depend on kind of lattice
    - extreme example !
        - a triangular lattice has more foldable structures

Are we finished with lattices ?