Secondary Structure Assignment

I619 Structural Bioinformatics

Introduction

- **Why assign secondary structure?**
  - the number of solved structures is large for manual assignment
  - there is disagreement among structural biologists as to what the best possible assignment is
- **Solution**
  - leave secondary structure assignment to the group who crystallized the protein
  - use automated methods and live with errors
  - the need for automated methods was recognized very early (1970s, 1980s)
- **DSSP program (1983)**
  - Define Secondary Structure of Proteins
- Secondary structure assignment starts with 3D structure and “defines” what are helices, sheets and coils. This is not prediction of secondary structure.

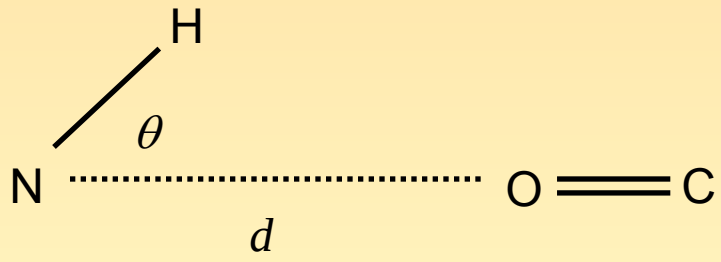
DSSP Algorithm

- Simple algorithm, physically motivated
- Based on H-bond patterns
  - concept of *n*-turn: an H-bond between CO of residue *i* and NH of residue *i* + *n* defines, where *n* = 3, 4, 5.
  - concept of bridge: an H-bond between CO and NH that are not near each other in the sequence

- **IDEA**
  - first find H-bonds
  - based on H-bonds, find turns and bridges
  - based on turns find α-helices, β-sheets, but accomodate imperfections
  - use geometric features and define bends, chirality, SS bonds and solvent exposure

Hydrogen Bond (H-bond)

- There is little wave overlap for H-bonded atoms, thus electrostatic model is precise



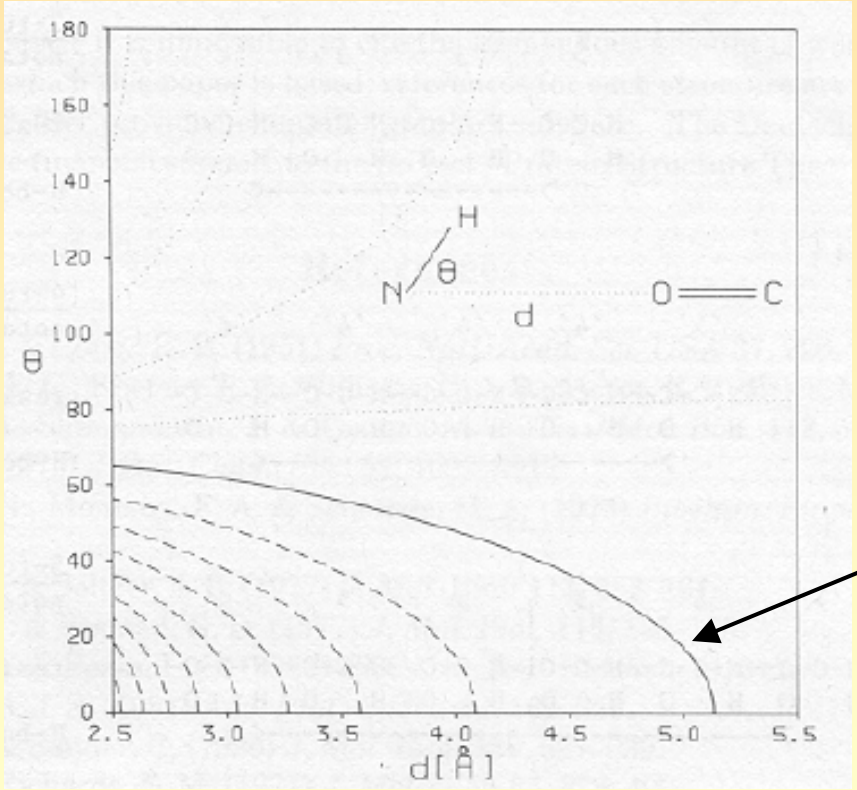
- Ideal bond length is for d = 2.9Å and θ = 0.

$$E = q_1 q_2 \cdot \left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}}\right) \cdot f$$

- f = 332 to obtain E in kcal/mol
- q_1 = 0.42e; q_2 = 0.20e, where e is charge of 1 electron; e = $1.6 \cdot 10^{-19}$ C

Hydrogen Bond (H-bond)

- There is little wave overlap for H-bonded atoms, thus electrostatic model is precise. There is no generally correct definition of H-bond.

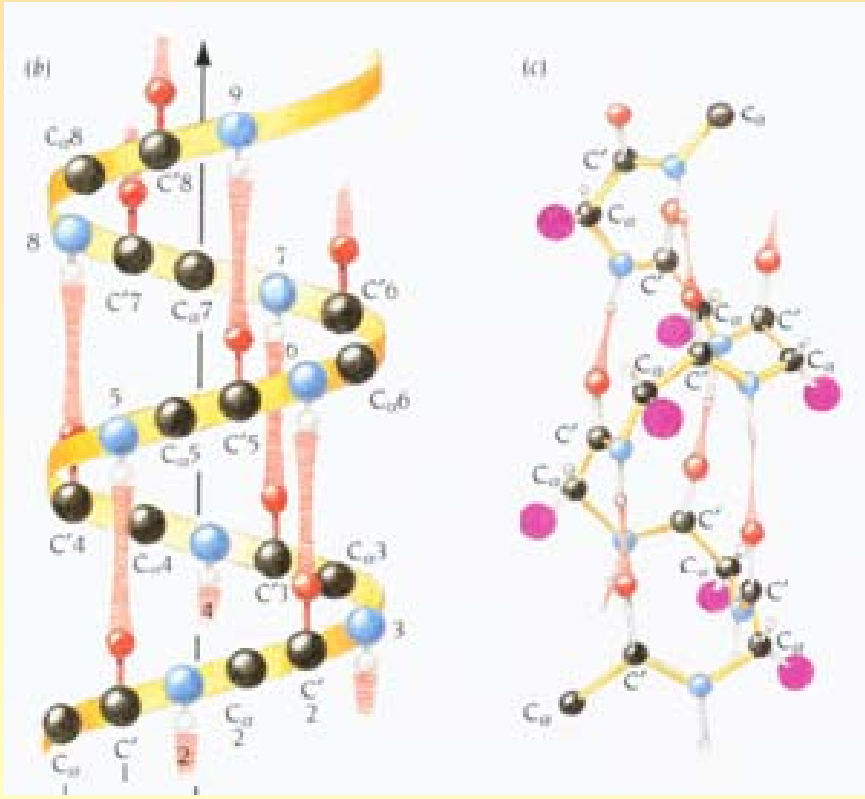


E = -0.5 kcal/mol

- *d* up to 5.2Å is allowed for θ = 0, and for optimal *d*, θ can go up to 63˚

n-Turns

- n-turn is assigned at residue *i* of there is an H-bond between CO(*i*) and NH(*i* + *n*), where *n* = 3, 4, 5.
- notation: n-turn = Hbond(*i*, *i* + *n*), *n* = 3, 4, 5.
- there exist 3-turn, 4-turn and 5-turn



Bridges

- Parallel:

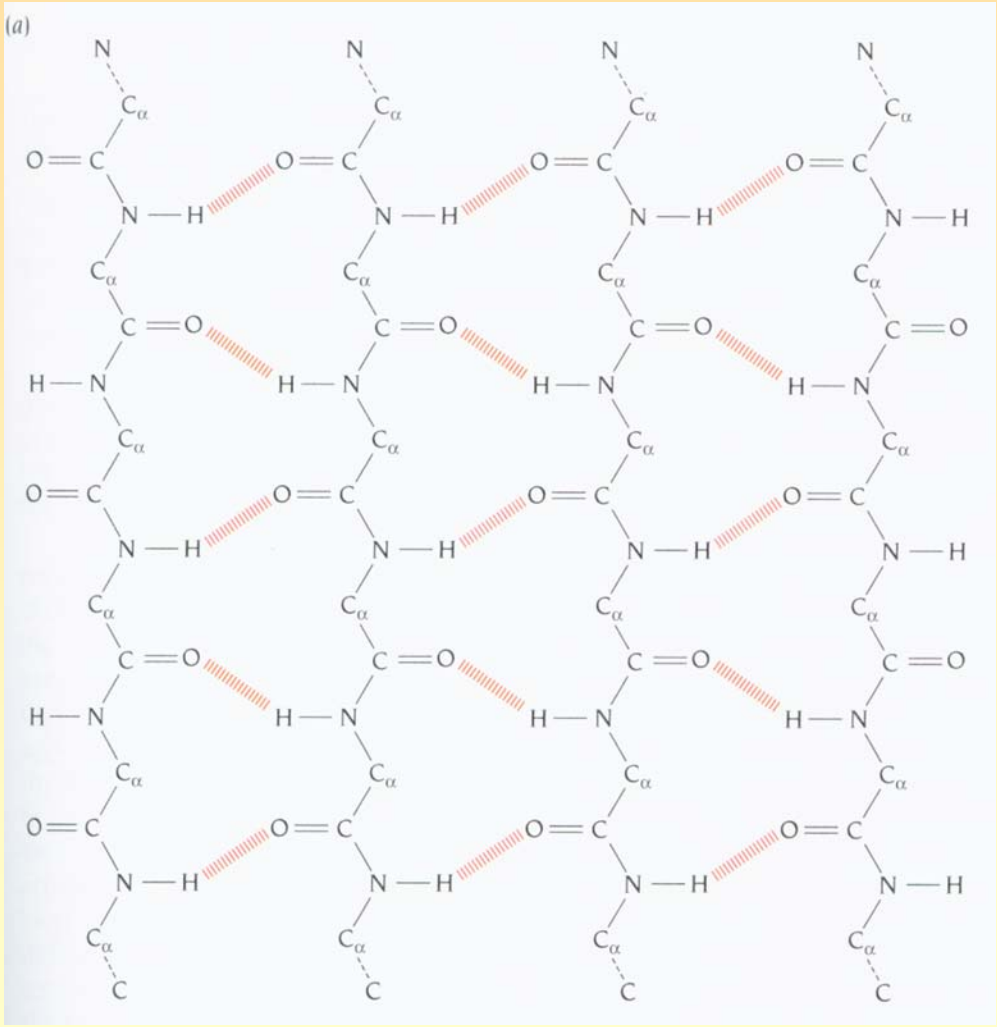
  Hbond(i - 1, j) and Hbond(j, i + 1) or

  Hbond(j - 1, i) and Hbond(i, j + 1)

- Antiparallel

  Hbond(i, j) and Hbond(j, i) or

  Hbond(i - 1, j + 1) and Hbond(j - 1, i + 1)



Bridges

- Parallel:

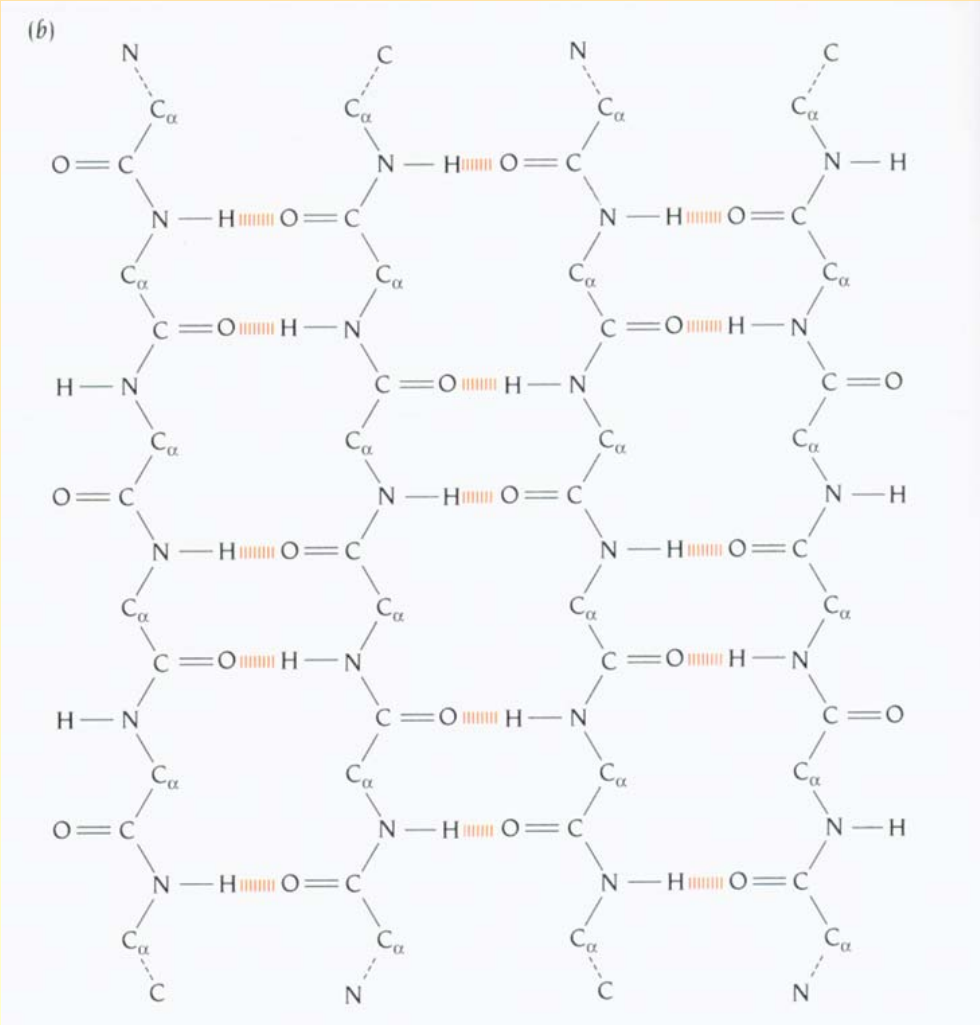
  Hbond(i - 1, j) and Hbond(j, i + 1) or

  Hbond(j - 1, i) and Hbond(i, j + 1)

- Antiparallel

  Hbond(i, j) and Hbond(j, i) or

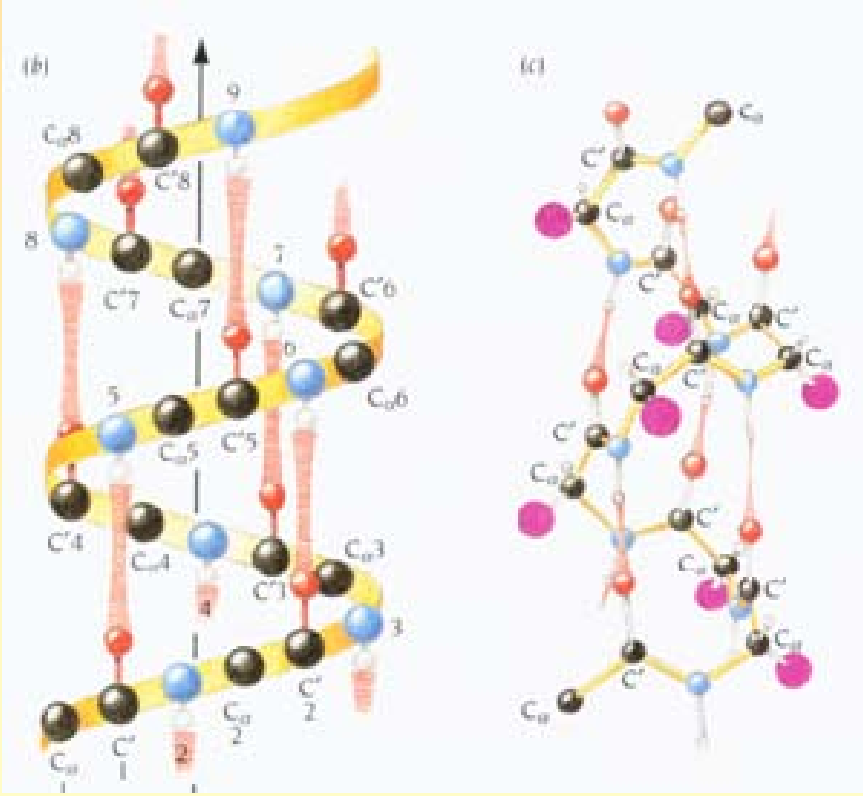
  Hbond(i - 1, j + 1) and Hbond(j - 1, i + 1)



Helices

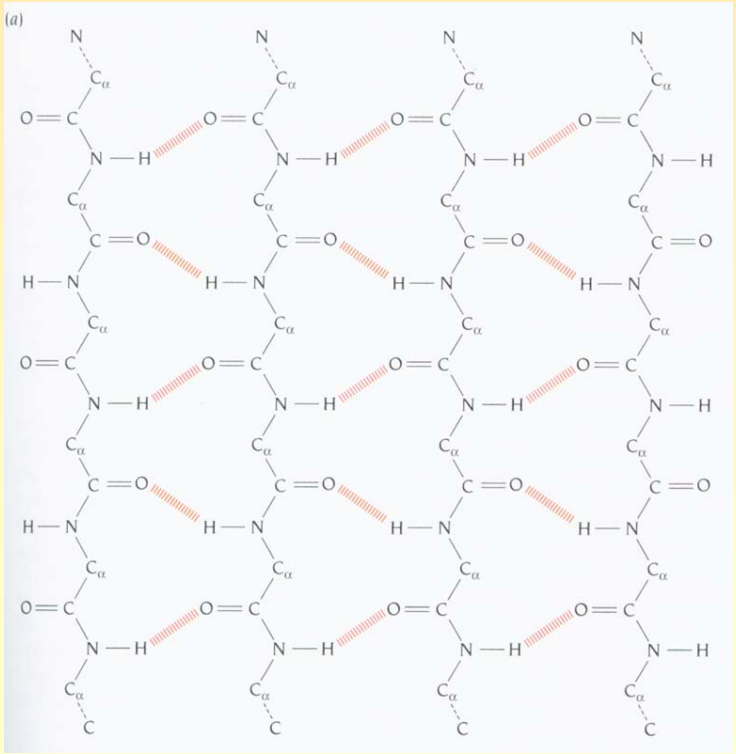
- Definition: A minimal helix has at least 2 n-turns
- Example: a 4-helix, of minimal length 4 from residues *i* to *i* + 3, requires 4-turns at residues *i* – 1 and *i*
-
- 4-helix(*i*, *i* + 3) = [4-turn(*i* – 1) and 4-turn(*i*)]

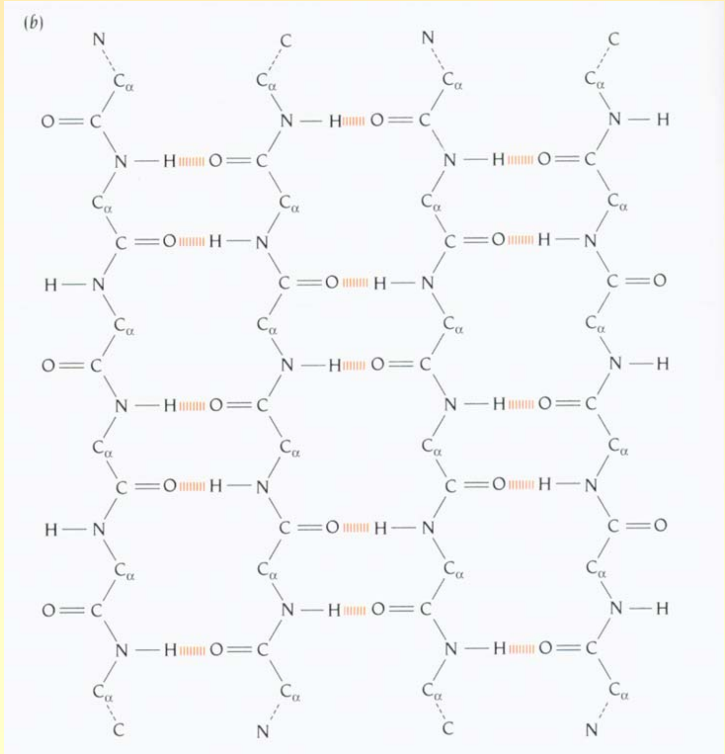
- longer helices are defined as overlaps of minimal helices
- one or two residue gaps are tolerated when connecting short helices to accommodate kinks



Sheets

- Definition:
  - ladder: set of one or more consecutive bridges of identical type
  - sheet: set of one or more ladders connected by shared residues





Geometrical Structures

- Bends:
  - regions with high curvature
  - chain curvature is quantified in the following way
  - Bend(i) = angle{$C^{\alpha}(i)$-$C^{\alpha}(i-2)$, $C^{\alpha}(i+2)$-$C^{\alpha}(i)$} > 70˚

- Chirality:
  - $\alpha(i)$ = dihedral angle {$C^{\alpha}(i-1)$, $C^{\alpha}(i)$, $C^{\alpha}(i+1)$-$C^{\alpha}(i+2)$}
  - DSSP reports only sign of α ("+" if 0˚ < α < 180˚ and "–" if -180˚ < α < 0˚)
  - most helices have positive chirality, most twisted ladders have negative chirality

3-D view of a left handed helix

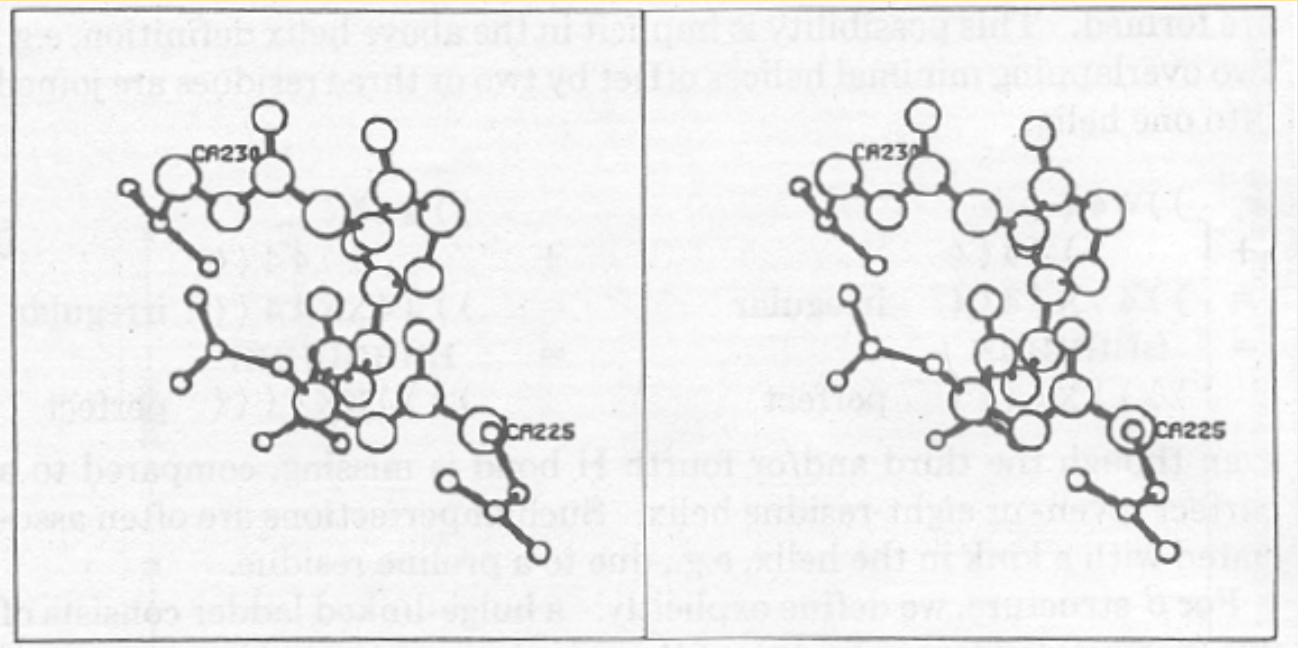


Fig. 5. Stereoviews of secondary structure: illustration of chirality. This short left-handed α-helix, Gln225–Val230 from thermolysin (2TLN) is the only one known to us. In Table AIII (note that chirality is entered at the second residue of each quartet) it appears as:

Geometrical Structures

- SS bonds:
  - covalent links between sulfurs of Cys residues
  - taken directly from PDB SSBOND records
  - S-S distance of 3Å is also a good definition

- Chain breaks:
  - if a peptide bond (C'-N) length exceeds 2.5Å this is considered to be a chain break
  - missing electron density or coordinate errors will determine chain breaks

Solvent Exposure

- The number of water molecules in direct contact with the protein or with particular part of protein

- Kabsch and Sander calculate this surface by integrating a step function *f* over all points *x* on the surface of a sphere of radius

  $r(atom) + r(water)$

  around atom *i*. *f* = 1 if a water molecule (approximated by sphere) centered at *x* (by definition in contact with atom *i*) does not intersect with any other protein atom.
- $r(O) = 1.4Å$; $r(N) = 1.65Å$; $r(C^{\alpha}) = 1.76Å$; $r(C) = 1.8Å$; $r(H_2O) = 1.4Å$

$$W = \frac{\text{Area}}{V(\text{water molecule})^{2/3}} \approx \frac{\text{Area}}{10}$$

W is the number of water molecules