Extracting information from published crystallographic data

Steps in solving a structure

clone gene

express protein

purify

#1

crystallize

#2, #3

measure X-ray diffraction

BMB695a

calculate electron density

build atomic model, interpret and publish

#4

Topics

Case 1: the refolding loop

- Missing atoms, B-factors and multiple crystal forms as a measure of flexibility
- What is the role of crystal contacts?
- Looking at electron density

Case 2: How do we get from A to B?

- Superimposing similar structures
- Looking for clashes

Case 3: How careful was I in 1999?

- Checking your favorite part of a protein structure for possible alternative interpretations

Case 1: from Selase's PhD defense

- Proteolytic assay to probe for transition between transcription initiation and elongation in T7 RNA polymerase
- Protein conformation of both initiation and elongation state are known (1QLN, 1MSW)
- Question from the audience: **"Does the refolding loop really refold, or does it transition from unfolded to folded?"**

Background info (straight from the thesis)

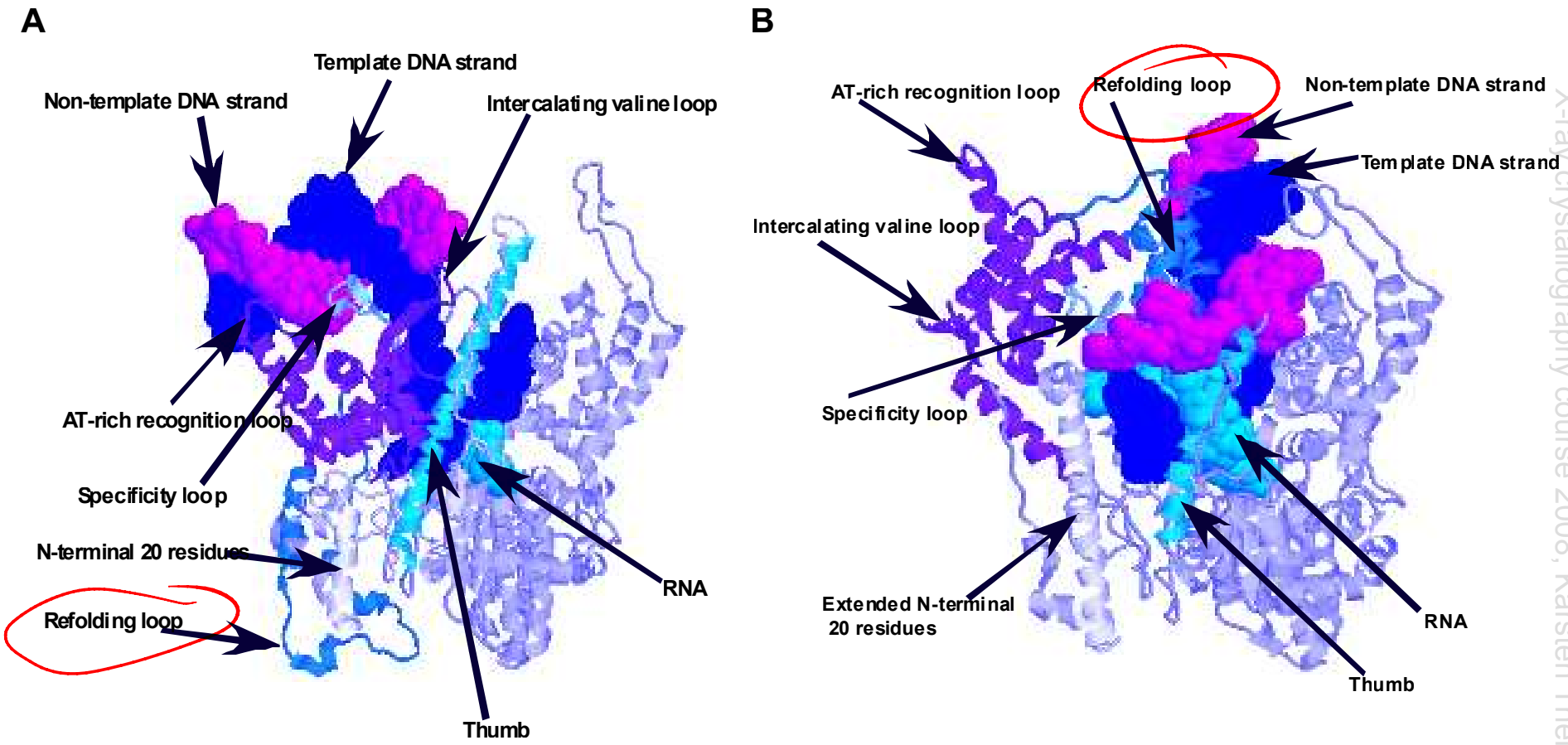


Figure 1.1.0: Labeled regions of the crystal structures of T7 RNA polymerase initiation and elongation complexes. (A) Initiation complex (1QLN); (B) Elongation complex (1MSW); Part of the C-terminal domain (residues 72-152 and 204-258) is color coded in grey, while the rigid N-terminal domain is colored forest green. In the elongation complex, the N-terminal 20 residues gets extended by residues 44-60 to 38 residues and the entire helix is colored yellow.

The proteolytic assay

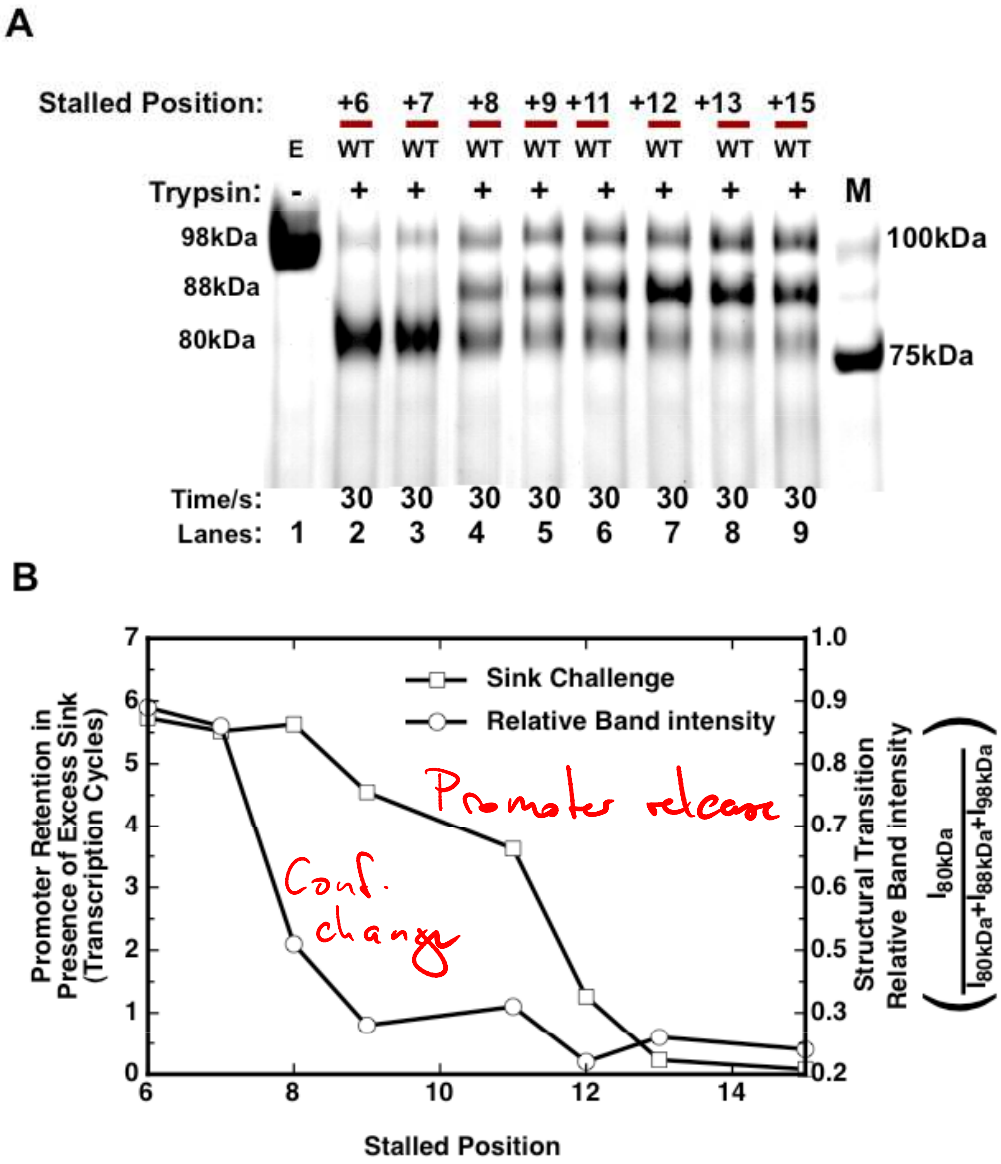


Figure 2.6.0: Partial tryptic digest of transcription complexes;
(A) SDS gel electrophoresis of complexes halted at the indicated position and then exposed to 1.2 equivalents of trypsin for 30 seconds prior to quenching in a 5X SDS gel loading buffer (0.225M Tris HCl, 5% SDS, 50% glycerol and 0.05% bromophenol blue) at room temperature. 'E' shows the polymerase only, without proteolysis; (B) Plot of sink challenge data from figure 2.4.0 (□) compared to plot of the relative band intensity ($I_{80kDa}/(I_{80kDa}+I_{88kDa}+I_{98kDa})$) at each stall position in "A" (○).

Case 1: list of tasks

1. Check for missing atoms/residues
2. Look at B-factors
3. Check for crystal contacts
4. Look at other crystal forms

The Protein Database (PDB)

- All journals require deposition of coordinates (sometimes instantly, sometimes after 6 mo)
- Not all journals require deposition of raw data

- Some information is easily available directly from the PDB, but some questions require downloading the crystallographic model and looking at it with specialized tools
- T7 polymerase during initiation: 1QLN

Tools

- Firstglance in Jmol
- Electron density server (only if raw data available)
  - Download density
  - How well do atoms fit to map?
  - Many other aspects
- What if? (you can upload your own coordinates)
  - Crystal contacts to 5 Å distance
  - Atomic clashes
- Helixweb, NIH
  - B-factor plot

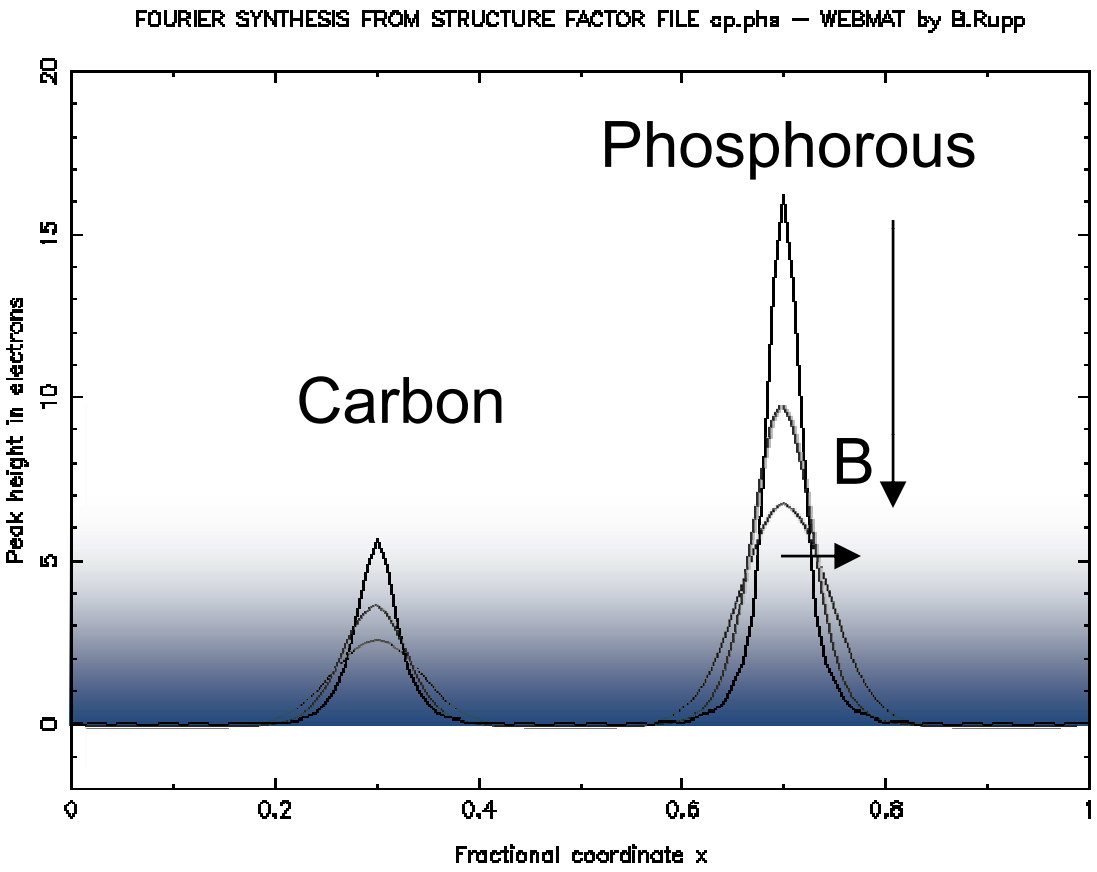
B-factors and disorder

- Synonyms: Temperature factors, atomic displacement factors
- B-factors describe how the electron density of an atom is broadened by static and dynamic disorder in the crystal

Static disorder: distinct atomic positions in different unit cells of the crystal
Dynamic disorder: changes in conformation over time during the measurement

B-factor	Displacement
20 Å^2	0.25 Å
40 Å^2	0.51 Å
80 Å^2	1.01 Å

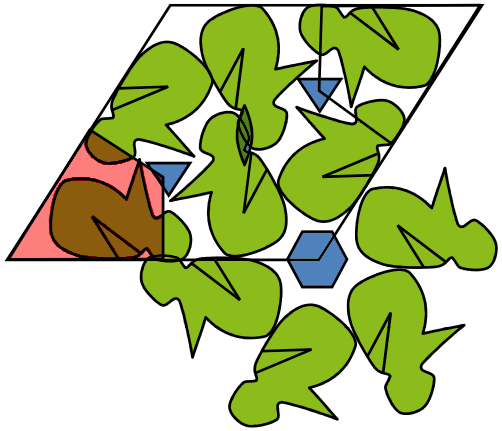
Disorder makes interpretation more difficult

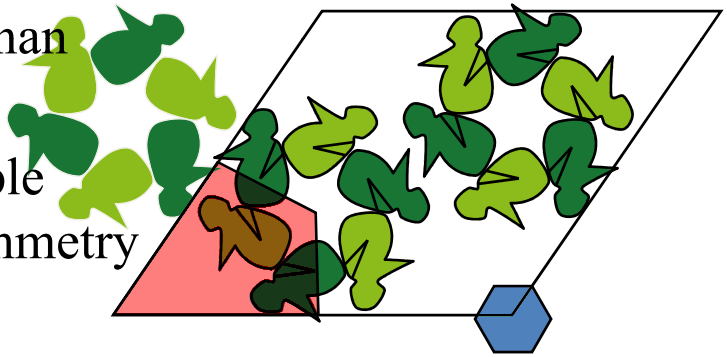


Shading represents noise due to
1) errors in experimental data
2) errors in phases derived from model

The asymmetric unit

- Crystal symmetry duplicates/triplicates etc. molecules placed into the unit cell
- The unique volume of the unit cell not related by crystal symmetry is called asymmetric unit
- Non-crystallographic symmetry (NCS)
  - the asymmetric unit often contains more than one copy of the protein
  - the operations superimposing these multiple copies are called non-crystallographic symmetry operations
  - NCS-related molecules aren't identical and have different crystal environments





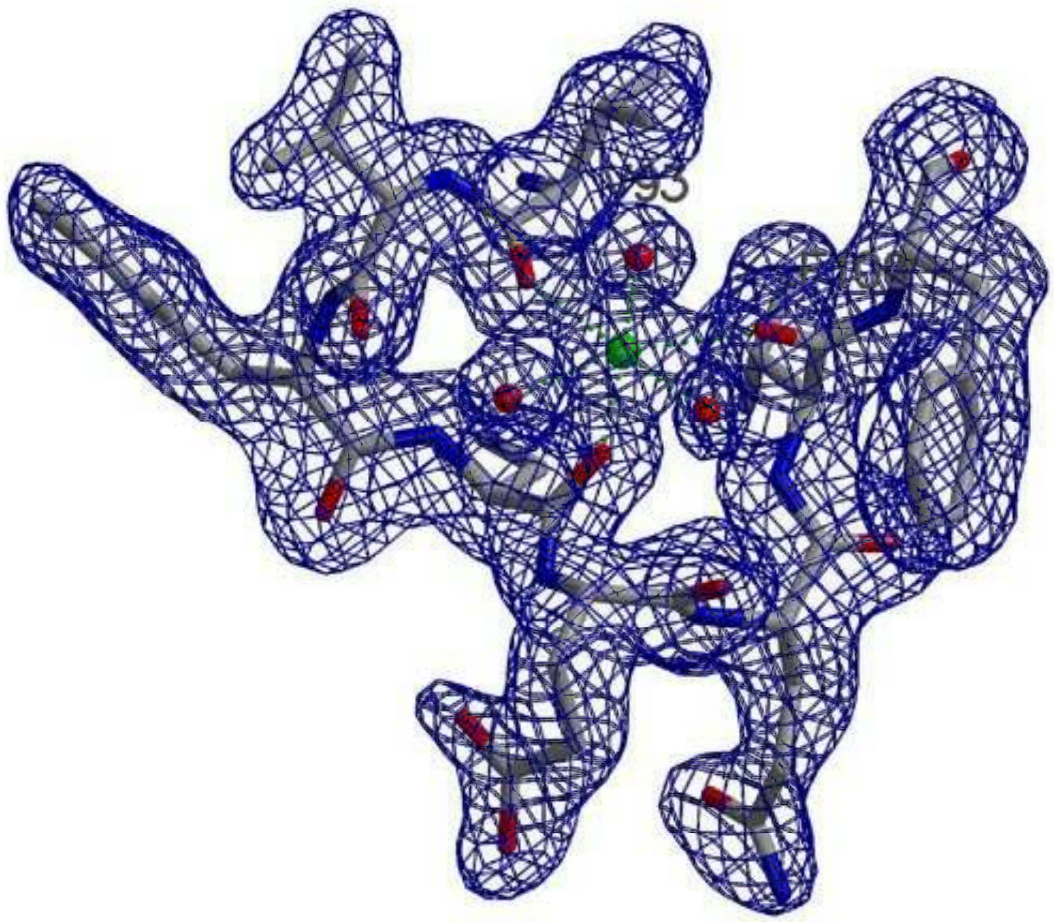
List of structures (from Uniprot)

1ARO,	X-ray, 2.80 Å,	P=1-883.
1CEZ,	X-ray, 2.40 Å,	A=1-883.
1H38,	X-ray, 2.90 Å,	A/B/C/D=1-883.
1MSW,	X-ray, 2.10 Å,	D=1-883.
1QLN,	X-ray, 2.40 Å,	A=1-883.
1S0V,	X-ray, 3.20 Å,	A/B/C/D=1-883.
1S76,	X-ray, 2.88 Å,	D=1-883.
1S77,	X-ray, 2.69 Å,	D=1-883.
2PI4,	X-ray, 2.50 Å,	A=6-883.
2PI5,	X-ray, 2.90 Å,	A=6-883.
4RNP,	X-ray, 3.00 Å,	A/B/C=1-883.

Case 2

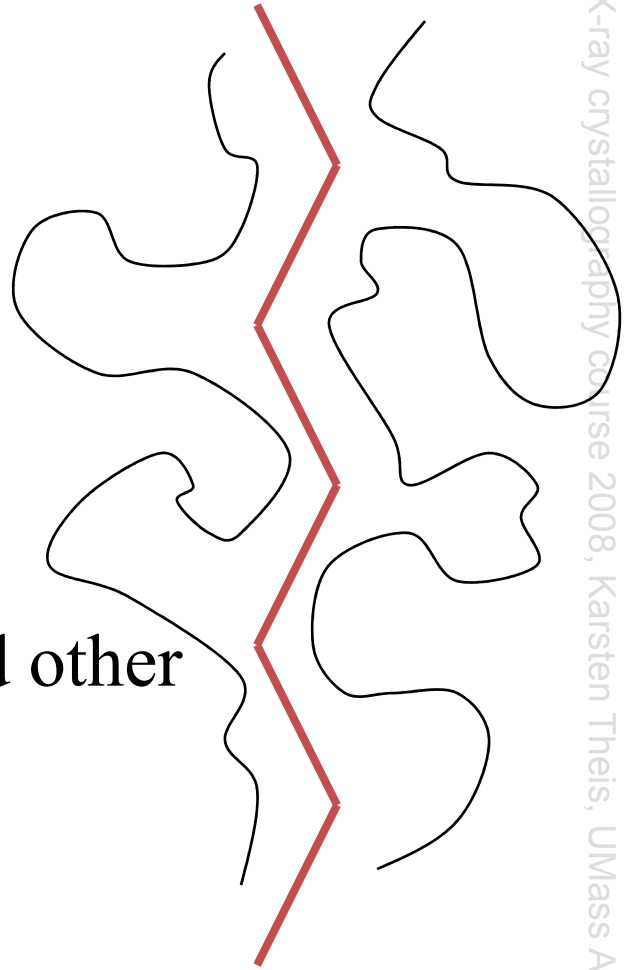
- Selase found that the refolding loop is protected from proteolysis before the promoter falls off
- Can we build a model in which the refolding loop is in the elongation conformation while the rest of the N-terminal part of T7 RNAP is still in its initiation conformation?
- Tasks
  - Superimpose the two structures
  - Look for clashes
  - Modify the resulting model to remove clashes

Interpretation of electron density



Building a model into the electron density involves interpretation and prior knowledge

- Protein/solvent regions
- C-alpha trace
- main chain, peptide direction
- sequence assignment
- side chain conformations
- disulfides, metals, glycosylation and other surprises



Pop quiz: interpreting electron density

1) Which pairs of amino acids have very similar electron density and are thus difficult to distinguish crystallographically? Asp/Glu ; Thr/Val ; Leu/Ile ; Lys/Met ; Asp/Asn ; Leu/Asp ; Glu/Gln
2) Which amino acids other than histidine have two side chain conformations resulting in almost identical electron density? What could help to distinguish the two possible conformations?