

# Bioinformatika III

## Trimačių struktūrų analizė ir spėjimas

Paskaita 12  
baltymų struktūrų tikrinimas  
(validacija)

Saulius Gražulis  
2008 m.

# Struktūrų nustatymo problemos

- Tiek eksperimentinės, tiek teoriškai nuspėtos struktūros gali turėti klaidų...
  - žmogiškų klaidų
  - programų/skaičiavimo klaidų
  - eksperimentinių klaidų
- Struktūros nustatomos/nuspėjamos baigtiniu tikslumu

# Atsitiktinės ir sisteminės paklaidos (Accuracy and precision)

- Mažos sisteminės paklaidos (high accuracy)
  - $\pi = 3.1$
- Mažos atsitiktinės paklaidos (high precision)
  - $\pi = 4.7352516121262$   
a very precise, but hopelessly inaccurate approximation...

# Klaidų tipai

- visiškai neteisingas modelis (foldas, sankloda)
- neteisingai pravesta grandinė
- „registro postūmis“
- neteisingos liekanos („taškinės mutacijos“)
- Neteisingos šoninės grandinės, neteisinga geometrija
- Netikslūs/neteisingi vanduo ir jonai

# Klaidų priežastys

- Kvalifikacijos stoka
- Skubėjimas, nepakankamas tarpinių žingsnių/modelių tikrinimas
- **Programos naudojamos kaip „juodos dėžės“, nesigilinant į jų veikimą**
- **Kokybės rodikliai stipriai koreliuoja su patikslinamais parametrais**

# Kas yra geras modelis?

- chemiškai priimtinas: teisingi jungčių ilgiai, kampai...
- fiziškai priimtinas: atomai „neužlipa“ vienas ant kito
- kristalografiškai teisingas: modelis neprieštarauja eksp. duomenims
- statistiškai priimtinas: modelis yra geriausia hipotezė stebėjimams paaiškinti
- struktūriškai priimtinas: panašus į kitas baltymų struktūras (Ramačandrano žemėlapiu, supakavimu)
- biologiškai priimtinas: suderinamas su žinoma biochemine ir biologine informacija

# Kas yra geras modelis?

- chemical sense: normal bond lengths and angles...
- physical sense: no interpenetrating non-bonded atoms...
- crystallographic sense: the model adequately explains the experimental data
- statistical sense: the model is the best hypothesis to explain the data...
- protein-structural sense: the model "looks like a protein": reasonable Ramachandran ...
- biological sense: the model explains observations, e.g. with respect to activity

# Validavimo kriterijai

- Geometriniai kriterijai: nukrypimai nuo idealių jungčių, chiraliskumas... (silpnas)
- Omega-kampas (silpnas)
- Phi-psi kampai, Ramačandrano diagrama (puikus)
- Chi1-chi2 kampai (vidutiniškas)
- Rotamerų konformacijos (puikus)



# Validavimo kriterijai (2)

- Ca-Ca parametrai (atstumai, kampai) (geras)
- Atomų kontaktų analizė (Directional atomic contact analysis, DACA) (puikus)
- Nekristalografinė simetrija (vidutiniškas)
- Temperatūriniai faktoriai (blogas)
- Duomenų skiriamoji geba (vidutiniškas)

# Validavimo kriterijai (3)

- Kristalografnis R-faktorius (dažniausiai blogas)
- „Laisvasis“ R-faktorius (puikus)
- $R_{\text{free}} - R_{\text{cryst}}$  (puikus)
- Tiesioginės erdvės R-faktorius ir koreliacijos koeficientas (geras, jei el. tankis neturi modelio įtakos)
- Spejamos koordinatų paklaidos (vidutiniškas)

# Validavimo įrankiai

- WHAT IF
- PROCHECK
- MOLEMAN2
- MolProbity
- Prosa
- Verify3D

# Validavimo kriterijų apžvalga

Nr.	Parameter	Validation potential
1.	The peptide phi and psi torsion angles (Ramachandran plot)	<i>excellent</i>
2.	Preferred rotamer conformations	<i>excellent</i>
3.	CA-CA distances ( $\sim 2.9$ Å for a <i>cis</i> -peptide, and $\sim 3.8$ Å for a <i>trans</i> -peptide), CA-CA-CA pseudo-angles, and CA-CA-CA-CA pseudo-torsions	<i>good</i>
4.	Directional atomic contact analysis (DACA): hydrophobic, electrostatic, and hydrogen-bonding interactions stabilising the protein structure	<i>excellent</i>
5.	The free R-value ( $R_{\text{free}}$ )	<i>excellent</i>
6.	$R_{\text{free}} - R$	<i>good</i>
7.	Real-space R-value: $\text{RSR} = \frac{\text{SUM}   \rho_{\text{obs}} - \rho_{\text{calc}}  }{\text{SUM}   \rho_{\text{obs}} + \rho_{\text{calc}}  }$ RSCC	<i>good.</i>  <i>(Provided the experimental map is not biased too much by the model.)</i>

# Validavimo kriterijų apžvalga (2)

Nr.	Parameter	Validation potential
1.	Bond lengths (the RMS deviation from ideal values); deviations from ideality of bond angles; chirality and planarity tests.	<i>poor</i>
2.	The peptide omega angle	<i>poor</i>
3.	Chi-1 (N-CA-CB-XG), chi-2 (CA-CB-XG-XD)	<i>moderate</i>
4.	Non-crystallographic symmetry	<i>moderate</i> <i>NCS constraints and restraints are so powerful that it is usually better to impose them during refinement (especially at low resolution) than to use NCS as an a posteriori validation criterion</i>
5.	Temperature factors, or B-factors	<i>poor</i>
6.	Resolution of the experimental data	<i>moderate</i>
7.	The crystallographic R-value: $R = \frac{\sum \text{weight} *    F_{\text{obs}}  - \text{scale} *  F_{\text{calc}}    }{\sum \text{weight} *  F_{\text{obs}}  }$	<i>poor</i>
8.	Estimated errors in the atomic coordinates and, by extension, in the atomic positions, bond lengths, etc.	<i>moderate if cross-validated, otherwise poor</i>