Homology Modeling

Maria Kalemanov

2013-06-04

Outline

Introduction

Protein structure modeling

- Goal: predict structure from sequence with accuracy comparable to best experimental results
- Major approaches to 3D structure prediction:
  - **Template based modeling**: template available
    - homology modeling (comparative modeling)
    - threading (fold recognition)
  - **Free modeling**: no template available
    - ab initio folding

Introduction

Protein structure modeling

- Goal: predict structure from sequence with accuracy comparable to best experimental results
- Major approaches to 3D structure prediction:
  - **Template based modeling**: template available
    - homology modeling (comparative modeling)
    - threading (fold recognition)
  - **Free modeling**: no template available
    - ab initio folding

Introduction

Protein structure modeling

- Goal: predict structure from sequence with accuracy comparable to best experimental results
- Major approaches to 3D structure prediction:
  - **Template based modeling**: template available
    - ★ homology modeling (comparative modeling)
    - ★ threading (fold recognition)
  - **Free modeling**: no template available
    - ★ ab initio folding

Introduction

Protein structure modeling

- Goal: predict structure from sequence with accuracy comparable to best experimental results
- Major approaches to 3D structure prediction:
  - **Template based modeling**: template available
    - homology modeling (comparative modeling)
    - threading (fold recognition)
  - **Free modeling**: no template available
    - ab initio folding

Introduction

Protein structure modeling

- Goal: predict structure from sequence with accuracy comparable to best experimental results
- Major approaches to 3D structure prediction:
  - **Template-based modeling (TBM)**:
    - ★ homology modeling (comparative modeling)
    - ★ threading (fold recognition)
  - **Template-free modeling (FM)**:
    - ★ ab initio folding

Introduction

Homology modeling

- Fundamental observations:
  - amino acid sequence $\rightarrow$ unique structure
  - evolution: structure changes much slower than sequence
    - similar sequences $\rightarrow$ almost identical structures
    - distantly related sequences $\rightarrow$ similar structures
    - two sequences in "safe homology modeling zone" have a similar structure

Introduction

Homology modeling

- Fundamental observations:
  - amino acid sequence $\rightarrow$ unique structure
  - evolution: structure changes much slower than sequence
    - similar sequences $\rightarrow$ almost identical structures
    - distantly related sequences $\rightarrow$ similar structures
    - two sequences in "safe homology modeling zone" have a similar structure

Introduction

Homology modeling

- Fundamental observations:
  - amino acid sequence $\rightarrow$ unique structure
  - evolution: structure changes much slower than sequence
    - similar sequences $\rightarrow$ almost identical structures
    - distantly related sequences $\rightarrow$ similar structures
    - two sequences in "safe homology modeling zone" have a similar structure

Introduction

Homology modeling

- Fundamental observations:
  - amino acid sequence $\rightarrow$ unique structure
  - evolution: structure changes much slower than sequence
    - similar sequences $\rightarrow$ almost identical structures
    - distantly related sequences $\rightarrow$ similar structures
    - two sequences in "safe homology modeling zone" have a similar structure

Introduction

Homology modeling

- Fundamental observations:
  - amino acid sequence $\rightarrow$ unique structure
  - evolution: structure changes much slower than sequence
    - similar sequences $\rightarrow$ almost identical structures
    - distantly related sequences $\rightarrow$ similar structures
    - two sequences in "safe homology modeling zone" have a similar structure

Introduction

Homology modeling

- Fundamental observations:
  - amino acid sequence $\rightarrow$ unique structure
  - evolution: structure changes much slower than sequence
    - ★ similar sequences $\rightarrow$ almost identical structures
    - ★ distantly related sequences $\rightarrow$ similar structures
    - ★ two sequences in "safe homology modeling zone" have a similar structure

Safe homology modeling zone

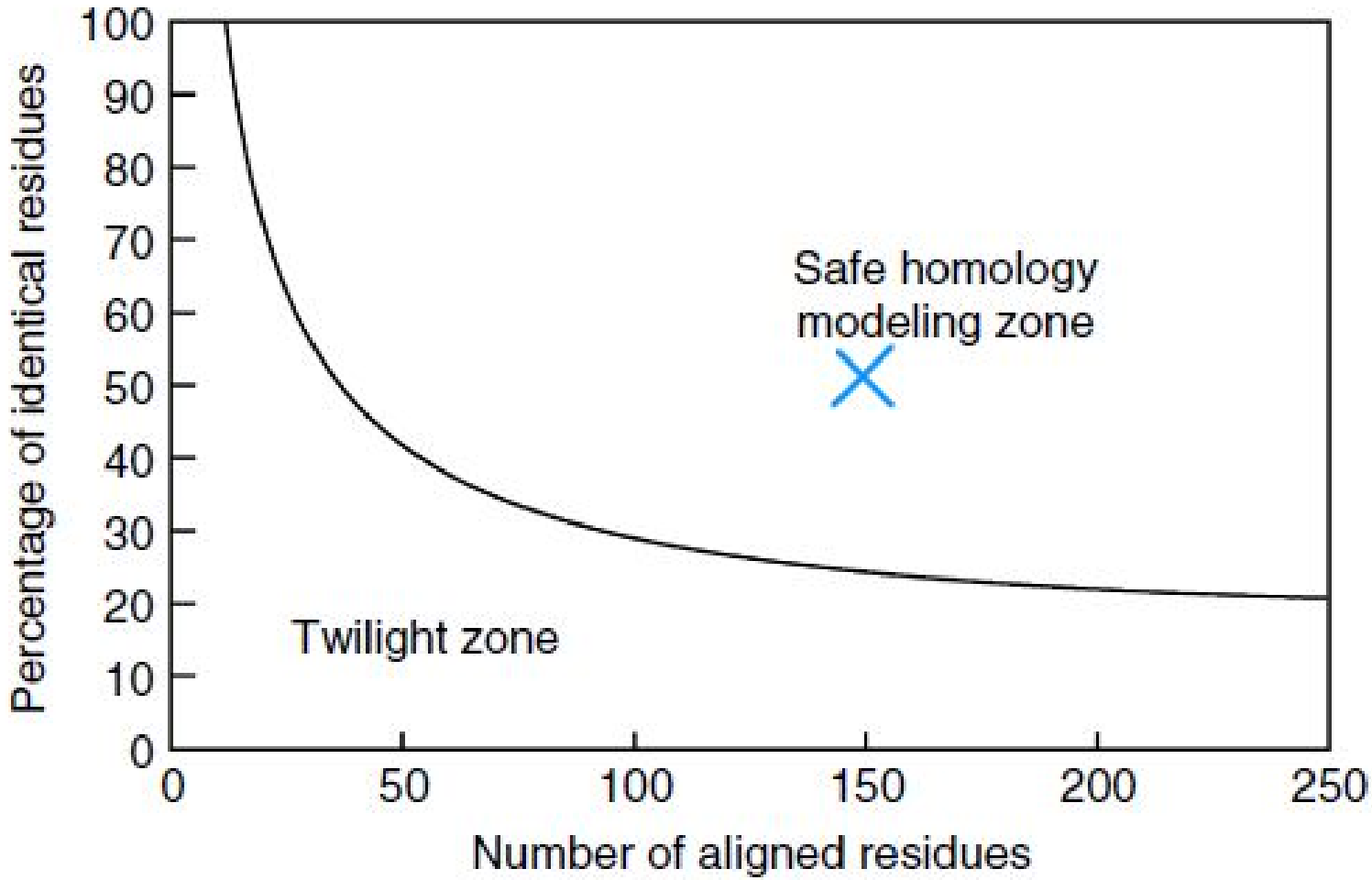


Figure: The two zones of sequence alignments. (Rost B 1999, Twilight zone of protein sequence alignments.)

Introduction

Homology modeling

- Terms
  - **Target**: the protein whose structure is to be modeled
  - **Template**: a homologue protein with known structure
- Length and % sequence identity of the alignment between the two must fall into the ”safe” region

Introduction

Homology modeling

- Terms
  - **Target**: the protein whose structure is to be modeled
  - **Template**: a homologue protein with known structure
- Length and % sequence identity of the alignment between the two must fall into the "safe" region

Introduction

Homology modeling

- Terms
  - **Target**: the protein whose structure is to be modeled
  - **Template**: a homologue protein with known structure
- Length and % sequence identity of the alignment between the two must fall into the "safe" region

Introduction

Homology modeling

- Terms
  - **Target**: the protein whose structure is to be modeled
  - **Template**: a homologue protein with known structure
- Length and % sequence identity of the alignment between the two must fall into the "safe" region

Introduction

Homology modeling

- The steps of homology modeling:
  1. Template selection and original alignment
  2. Alignment correction
  3. Backbone generation
  4. Loop modeling
  5. Side-chain modeling
  6. Model optimization
  7. Model validation
- Steps 3-6: model building

Introduction

Homology modeling

- The steps of homology modeling:
  1. Template selection and original alignment
  2. Alignment correction
  3. Backbone generation
  4. Loop modeling
  5. Side-chain modeling
  6. Model optimization
  7. Model validation
- Steps 3-6: model building

Introduction

Homology modeling

- The steps of homology modeling:
  1. Template selection and original alignment
  2. Alignment correction
  3. Backbone generation
  4. Loop modeling
  5. Side-chain modeling
  6. Model optimization
  7. Model validation
- Steps 3-6: model building

Homology modeling steps

1. Template selection and original alignment

- Use sequence homology search and alignment tool, e.g. BLAST, with:
  - query sequence: target
  - database: PDB
- Obtained list of hits - the templates and alignments

Homology modeling steps

1. Template selection and original alignment

- Use sequence homology search and alignment tool, e.g. BLAST, with:
  - query sequence: target
  - database: PDB
- Obtained list of hits - the templates and alignments

Homology modeling steps

1. Template selection and original alignment

- Use sequence homology search and alignment tool, e.g. BLAST, with:
  - query sequence: target
  - database: PDB
- Obtained list of hits - the templates and alignments

Homology modeling steps

1. Template selection and original alignment

- Use sequence homology search and alignment tool, e.g. BLAST, with:
  - query sequence: target
  - database: PDB
- Obtained list of hits - the templates and alignments

Homology modeling steps

2. Alignment correction

- Region of low sequence identity $\rightarrow$ use MSA with other sequences

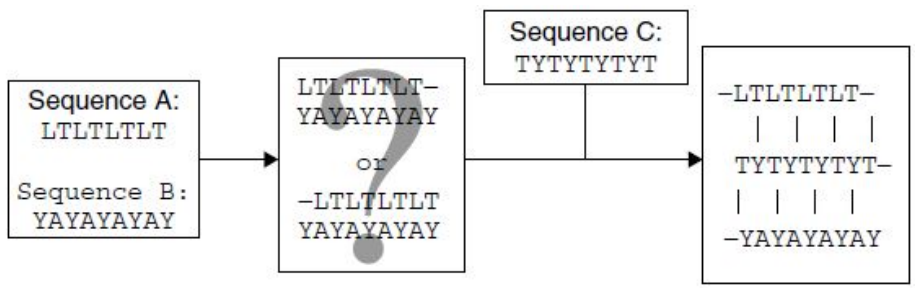


Figure: A third sequence C enables to align sequences A and B
(Krieger et al., Homology Modeling)

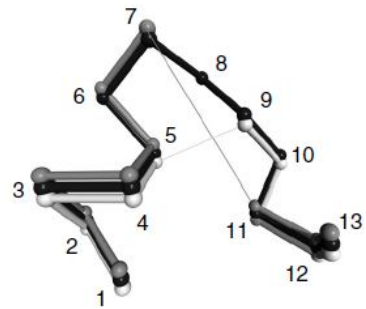
Homology modeling steps

2. Alignment correction

- MSA helps to place deletions or insertions

		1	2	3	4	5	6	7	8	9	10	11	12	13
Template		PHE	ASP	ILE	CYS	ARG	LEU	PRO	GLY	SER	ALA	GLU	ALA	VAL
Model(bad)	1	PHE	ASN	VAL	CYS	ARG	ALA	PRO	---	---	---	GLU	ALA	ILE
Model(good)	2	PHE	ASN	VAL	CYS	ARG	---	---	---	ALA	PRO	GLU	ALA	ILE

(a)



(b)

Figure: (a) The 3-residue deletion appears to be modeled better in the first alignment (aligned proline). (b) Structural alignment with template (C_α-trace: black). While the alignment with the higher score (dark gray) leads to a big gap of 7.5Å, the second alignment has only a tiny gap of 1.3Å. (Krieger et al., Homology Modeling)

Homology modeling steps

3. Backbone generation

- Copy coordinates of aligned residues from the template
  - if different residues $\rightarrow$ only copy the backbone coordinates (N, C_α, C and O)
  - if identical residues $\rightarrow$ side chain can also be included
- Multiple template modeling: if two templates align (well) to different regions $\rightarrow$ combine the good regions of both templates into one model

Homology modeling steps

3. Backbone generation

- Copy coordinates of aligned residues from the template
  - if different residues $\rightarrow$ only copy the backbone coordinates (N, C_α, C and O)
  - if identical residues $\rightarrow$ side chain can also be included
- Multiple template modeling: if two templates align (well) to different regions $\rightarrow$ combine the good regions of both templates into one model

Homology modeling steps

3. Backbone generation

- Copy coordinates of aligned residues from the template
  - if different residues $\rightarrow$ only copy the backbone coordinates (N, C_α, C and O)
  - if identical residues $\rightarrow$ side chain can also be included
- Multiple template modeling: if two templates align (well) to different regions $\rightarrow$ combine the good regions of both templates into one model

Homology modeling steps

3. Backbone generation

- Copy coordinates of aligned residues from the template
  - if different residues $\rightarrow$ only copy the backbone coordinates (N, C_α, C and O)
  - if identical residues $\rightarrow$ side chain can also be included
- Multiple template modeling: if two templates align (well) to different regions $\rightarrow$ combine the good regions of both templates into one model

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Treatment of gaps:
  - in the model (deletions) $\rightarrow$ omit residues from the template
  - in the template (insertions) $\rightarrow$ cut the template backbone, insert the missing residues
- $\rightarrow$ conformational change in backbone
  - cannot happen in secondary structure elements $\rightarrow$ in loops and turns
  - difficult to predict!
  - often different also in aligned loops!

Homology modeling steps

4. Loop modeling

- Main approaches to loops modeling:
  1. Knowlegde based
     - search PDB for known loops, copy the coordinates
  2. Energy based
     - minimize an energy function to get the best loop conformation

Homology modeling steps

4. Loop modeling

- Main approaches to loops modeling:
  1. Knowlegde based
     - ⋆ search PDB for known loops, copy the coordinates
  2. Energy based
     - ⋆ minimize an energy function to get the best loop conformation

Homology modeling steps

4. Loop modeling

- Main approaches to loops modeling:
  1. Knowlegde based
     - ⋆ search PDB for known loops, copy the coordinates
  2. Energy based
     - ⋆ minimize an energy function to get the best loop conformation

Homology modeling steps

5. Side-chain modeling

- Side-chain conformations (rotamers) are conserved in structures with high sequence identity
  $\rightarrow$ copy conserved residues entirely
- Otherwise: use libraries of common rotamers

Homology modeling steps

5. Side-chain modeling

- Side-chain conformations (rotamers) are conserved in structures with high sequence identity
  $\rightarrow$ copy conserved residues entirely
- Otherwise: use libraries of common rotamers

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Rotamer prediction and energy minimization
  1. predict the rotamers
  2. predict the shifts in the backbone
  3. predict the rotamers for the new backbone
  4. go to step 2
- repeat until convergence
- Strongly depends on energy function

Homology modeling steps

6. Model optimization

- Alternative: run a molecular dynamics simulation of the model
  - implicitly contains entropic effects
  - force fields are not precise enough!

Homology modeling steps

6. Model optimization

- Alternative: run a molecular dynamics simulation of the model
  - implicitly contains entropic effects
  - force fields are not precise enough!

Homology modeling steps

6. Model optimization

- Alternative: run a molecular dynamics simulation of the model
  - implicitly contains entropic effects
  - force fields are not precise enough!

Homology modeling steps

7. Model validation

- Number of errors mainly depends on:
  1. % sequence identity between template and target
     - >90% → accuracy compared to crystallographically determined structures
     - 50%-90% → RMSD error 1.5Å (larger local errors)
     - near 25% → very large error
  2. number of errors in the template

Homology modeling steps

7. Model validation

- Number of errors mainly depends on:
  1. % sequence identity between template and target
     - $>90\%$ $\rightarrow$ accuracy compared to crystallographically determined structures
     - 50%-90% $\rightarrow$ RMSD error 1.5Å (larger local errors)
     - near 25% $\rightarrow$ very large error
  2. number of errors in the template

Homology modeling steps

7. Model validation

- Number of errors mainly depends on:
  1. % sequence identity between template and target
     - ★ >90% → accuracy compared to crystallographically determined structures
     - ★ 50%-90% → RMSD error 1.5Å (larger local errors)
     - ★ near 25% → very large error
  2. number of errors in the template

Homology modeling steps

7. Model validation

- Number of errors mainly depends on:
  1. % sequence identity between template and target
     - ★ >90% → accuracy compared to crystallographically determined structures
     - ★ 50%-90% → RMSD error 1.5Å (larger local errors)
     - ★ near 25% → very large error
  2. number of errors in the template

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

Homology modeling steps

7. Model validation

Ways to estimate errors:

1. Calculate the model's force field energy
   - check bond length/angles, number of bumps
   - well-minimized models can still be misfolded!
   - one number for protein, not for single aa!
2. Determine of "normality indices", e.g.:
   - inside/outside distribution of polar/apolar residues
   - probability to find other atoms at a given distance from certain atom ("radial distribution function")
   - direction of atomic contacts ("3D distribution function")

SWISS-MODEL

User interaction modes

1. **first approach mode**
   - input: amino acid sequence of the target
   - optionally: up to 5 template structures
2. **alignment mode**
   - input: user-defined target-template alignment
3. **project mode**
   - input: project file from DeepView (Swiss-PdbViewer) with manually optimized target-template alignments
   - control over a wide range of parameters

SWISS-MODEL

User interaction modes

1. **first approach mode**
   - input: amino acid sequence of the target
   - optionally: up to 5 template structures
2. **alignment mode**
   - input: user-defined target-template alignment
3. **project mode**
   - input: project file from DeepView (Swiss-PdbViewer) with manually optimized target-template alignments
   - control over a wide range of parameters

SWISS-MODEL

User interaction modes

1. **first approach mode**
   - input: amino acid sequence of the target
   - optionally: up to 5 template structures
2. **alignment mode**
   - input: user-defined target-template alignment
3. **project mode**
   - input: project file from DeepView (Swiss-PdbViewer) with manually optimized target-template alignments
   - control over a wide range of parameters

SWISS-MODEL

Modeling procedure

- Four general steps:
  1. template selection
  2. target template alignment
  3. model building
  4. evaluation
- SWISS-MODEL approach: **rigid fragment assembly**
- The procedure steps can be iteratively repeated

SWISS-MODEL

Modeling procedure

- Four general steps:
  1. template selection
  2. target template alignment
  3. model building
  4. evaluation
- SWISS-MODEL approach: **rigid fragment assembly**
- The procedure steps can be iteratively repeated

SWISS-MODEL

1. Template selection

- Extraction of template structure library ExPDB from PDB
  - Splitting of PDB files into protein chains
  - Removement of unreliable entries
  - Addition of other useful information to the header
- Searching templates for a given protein in the library
- If found templates cover separate regions of the target $\rightarrow$ modeling separate batches independently

SWISS-MODEL

1. Template selection

- Extraction of template structure library ExPDB from PDB
  - Splitting of PDB files into protein chains
  - Removement of unreliable entries
  - Addition of other useful information to the header
- Searching templates for a given protein in the library
- If found templates cover separate regions of the target $\rightarrow$ modeling separate batches independently

SWISS-MODEL

1. Template selection

- Extraction of template structure library ExPDB from PDB
  - Splitting of PDB files into protein chains
  - Removement of unreliable entries
  - Addition of other useful information to the header
- Searching templates for a given protein in the library
- If found templates cover separate regions of the target $\rightarrow$ modeling separate batches independently

SWISS-MODEL

1. Template selection

- Extraction of template structure library ExPDB from PDB
  - Splitting of PDB files into protein chains
  - Removement of unreliable entries
  - Addition of other useful information to the header
- Searching templates for a given protein in the library
- If found templates cover separate regions of the target $\rightarrow$ modeling separate batches independently

SWISS-MODEL

1. Template selection

- Extraction of template structure library ExPDB from PDB
  - Splitting of PDB files into protein chains
  - Removement of unreliable entries
  - Addition of other useful information to the header
- Searching templates for a given protein in the library
- If found templates cover separate regions of the target $\rightarrow$ modeling separate batches independently

SWISS-MODEL

1. Template selection

- Extraction of template structure library ExPDB from PDB
  - Splitting of PDB files into protein chains
  - Removement of unreliable entries
  - Addition of other useful information to the header
- Searching templates for a given protein in the library
- If found templates cover separate regions of the target $\rightarrow$ modeling separate batches independently

SWISS-MODEL

2. Target template alignment

- Superposition of max. 5 templates per batch using iterative least squares algorithm
- Removement of templates with too high RMSD to the first template
- Generation of a structural alignment
- Calculation of local pair-wise alignments of the target to the templates
- Heuristic improvement of the alignment

SWISS-MODEL

2. Target template alignment

- Superposition of max. 5 templates per batch using iterative least squares algorithm
- Removement of templates with too high RMSD to the first template
- Generation of a structural alignment
- Calculation of local pair-wise alignments of the target to the templates
- Heuristic improvement of the alignment

SWISS-MODEL

2. Target template alignment

- Superposition of max. 5 templates per batch using iterative least squares algorithm
- Removement of templates with too high RMSD to the first template
- Generation of a structural alignment
- Calculation of local pair-wise alignments of the target to the templates
- Heuristic improvement of the alignment

SWISS-MODEL

2. Target template alignment

- Superposition of max. 5 templates per batch using iterative least squares algorithm
- Removement of templates with too high RMSD to the first template
- Generation of a structural alignment
- Calculation of local pair-wise alignments of the target to the templates
- Heuristic improvement of the alignment

SWISS-MODEL

2. Target template alignment

- Superposition of max. 5 templates per batch using iterative least squares algorithm
- Removement of templates with too high RMSD to the first template
- Generation of a structural alignment
- Calculation of local pair-wise alignments of the target to the templates
- Heuristic improvement of the alignment

SWISS-MODEL

3 Model building

- Generation of model core
  - Weighting of templates by sequence similarity to the target
  - Averaging the backbone atom positions of the templates
- Modeling loop regions using constraint space programming (CSP)
- If CSP solution is not good or loops > 10 residues $\rightarrow$ searching a loop library

SWISS-MODEL

3 Model building

- Generation of model core
  - Weighting of templates by sequence similarity to the target
  - Averaging the backbone atom positions of the templates
- Modeling loop regions using constraint space programming (CSP)
- If CSP solution is not good or loops > 10 residues $\rightarrow$ searching a loop library

SWISS-MODEL

3 Model building

- Generation of model core
  - Weighting of templates by sequence similarity to the target
  - Averaging the backbone atom positions of the templates
- Modeling loop regions using constraint space programming (CSP)
- If CSP solution is not good or loops > 10 residues $\rightarrow$ searching a loop library

SWISS-MODEL

3 Model building

- Generation of model core
  - Weighting of templates by sequence similarity to the target
  - Averaging the backbone atom positions of the templates
- Modeling loop regions using constraint space programming (CSP)
- If CSP solution is not good or loops > 10 residues $\rightarrow$ searching a loop library

SWISS-MODEL

3 Model building

- Generation of model core
  - Weighting of templates by sequence similarity to the target
  - Averaging the backbone atom positions of the templates
- Modeling loop regions using constraint space programming (CSP)
- If CSP solution is not good or loops > 10 residues $\rightarrow$ searching a loop library

SWISS-MODEL

3 Model building: side chain modeling

- Starting with conserved residues:
  construction of model side chains from the template
- Selection of the most likely side chain conformation
  from a backbone dependent rotamer library

SWISS-MODEL

3 Model building: side chain modeling

- Starting with conserved residues:
  construction of model side chains from the template
- Selection of the most likely side chain conformation
  from a backbone dependent rotamer library

- Joining of rigid fragments
  $\rightarrow$ deviations in target structure geometry
  $\rightarrow$ regulation by **steepest descent energy minimization**

SWISS-MODEL

4. Evaluation

- Several tools to evaluate reliability of the model:
  - **C-score** estimates variability of the template structure at each position (IN/DELs: C-score=99)
  - **Force field energy for the overall structure/each residue** identifies regions with conformational or electrostatic problems
  - **WhatCheck** report and evaluation by atomic mean force potential **ANOLEA** (optional)
- For alignment with sequence identity below 40% manual editing is necessary! $\rightarrow$ **DeepView**

SWISS-MODEL

4. Evaluation

- Several tools to evaluate reliability of the model:
  - **C-score** estimates variability of the template structure at each position (IN/DELs: C-score=99)
  - **Force field energy for the overall structure/each residue** identifies regions with conformational or electrostatic problems
  - **WhatCheck** report and evaluation by atomic mean force potential **ANOLEA** (optional)
- For alignment with sequence identity below 40% manual editing is necessary! $\rightarrow$ **DeepView**

SWISS-MODEL

4. Evaluation

- Several tools to evaluate reliability of the model:
  - **C-score** estimates variability of the template structure at each position (IN/DELs: C-score=99)
  - **Force field energy for the overall structure/each residue** identifies regions with conformational or electrostatic problems
  - **WhatCheck** report and evaluation by atomic mean force potential **ANOLEA** (optional)
- For alignment with sequence identity below 40% manual editing is necessary! $\rightarrow$ **DeepView**

SWISS-MODEL

4. Evaluation

- Several tools to evaluate reliability of the model:
  - **C-score** estimates variability of the template structure at each position (IN/DELs: C-score=99)
  - **Force field energy for the overall structure/each residue** identifies regions with conformational or electrostatic problems
  - **WhatCheck** report and evaluation by atomic mean force potential **ANOLEA** (optional)
- For alignment with sequence identity below 40% manual editing is necessary! → **DeepView**

SWISS-MODEL

4. Evaluation

- Several tools to evaluate reliability of the model:
  - **C-score** estimates variability of the template structure at each position (IN/DELs: C-score=99)
  - **Force field energy for the overall structure/each residue** identifies regions with conformational or electrostatic problems
  - **WhatCheck** report and evaluation by atomic mean force potential **ANOLEA** (optional)
- For alignment with sequence identity below 40% manual editing is necessary! $\rightarrow$ **DeepView**

I-TASSER

- Best fully automated structure prediction server in CASP8 (2008)
- Pipeline
  1. Template identification
  2. Structure assembly
  3. Atomic model construction
  4. Model selection

I-TASSER

- Best fully automated structure prediction server in CASP8 (2008)
- Pipeline
  1. Template identification
  2. Structure assembly
  3. Atomic model construction
  4. Model selection

I-TASSER

Pipeline

1. Template identification
   threading through a PDB structure library using
   - MUSTER
     profile-profile alignment algorithm
   - LOMETS
     local meta-threading server including HHSEARCH and 4 others

I-TASSER

Pipeline

1. Template identification
   threading through a PDB structure library using
   - MUSTER
     profile-profile alignment algorithm
   - LOMETS
     local meta-threading server including HHSEARCH and 4 others

I-TASSER

Pipeline

1. Template identification
   threading through a PDB structure library using
   - MUSTER
     profile-profile alignment algorithm
   - LOMETS
     local meta-threading server including HHSEARCH and 4 others

I-TASSER

Pipeline

2. Structure assembly
   - Continuous fragments: assembled into full-length models
   - Unaligned loops: *ab initio* modeling
   - Two sets of simulations with the following query structures:
     1. threading templates
     2. SPICKER cluster centroids (reduced models)

     $\rightarrow$ refine local geometry and global topology of the centroids

I-TASSER

Pipeline

2. Structure assembly
   - Continuous fragments: assembled into full-length models
   - Unaligned loops: *ab initio* modeling
   - Two sets of simulations with the following query structures:
     1. threading templates
     2. SPICKER cluster centroids (reduced models)

     → refine local geometry and global topology of the centroids

I-TASSER

Pipeline

2. Structure assembly
   - Continuous fragments: assembled into full-length models
   - Unaligned loops: *ab initio* modeling
   - Two sets of simulations with the following query structures:
     1. threading templates
     2. SPICKER cluster centroids (reduced models)

     → refine local geometry and global topology of the centroids

I-TASSER

Pipeline

2. Structure assembly
   - Continuous fragments: assembled into full-length models
   - Unaligned loops: *ab initio* modeling
   - Two sets of simulations with the following query structures:
     1. threading templates
     2. SPICKER cluster centroids (reduced models)

     $\rightarrow$ refine local geometry and global topology of the centroids

I-TASSER

Pipeline

2. Structure assembly
   - Continuous fragments: assembled into full-length models
   - Unaligned loops: *ab initio* modeling
   - Two sets of simulations with the following query structures:
     1. threading templates
     2. SPICKER cluster centroids (reduced models)

        $\rightarrow$ refine local geometry and global topology of the centroids

I-TASSER

Pipeline

2. Structure assembly
   - Continuous fragments: assembled into full-length models
   - Unaligned loops: *ab initio* modeling
   - Two sets of simulations with the following query structures:
     1. threading templates
     2. SPICKER cluster centroids (reduced models)
        $\rightarrow$ refine local geometry and global topology of the centroids

I-TASSER

Pipeline

2. Structure assembly
   - Knowledge-based force field:
     1. general statistic terms from PDB (C_α/side chain correlations, H-bonds, hydrophobicity)
     2. spatial restraints from threading templates
     3. sequence-based contact prediction from SVMSEQ
   - 9 sets of predictions:
     - C_α, C_β and side-chains centers
     - contact cutoffs: 6Å, 7Åand 8Å

I-TASSER

Pipeline

2. Structure assembly
   - Knowledge-based force field:
     1. general statistic terms from PDB (C_α/side chain correlations, H-bonds, hydrophobicity)
     2. spatial restraints from threading templates
     3. sequence-based contact prediction from SVMSEQ
   - 9 sets of predictions:
     - C_α, C_β and side-chains centers
     - contact cutoffs: 6Å, 7Åand 8Å

I-TASSER

Pipeline

2. Structure assembly
   - Knowledge-based force field:
     1. general statistic terms from PDB (C_α/side chain correlations, H-bonds, hydrophobicity)
     2. spatial restraints from threading templates
     3. sequence-based contact prediction from SVMSEQ
   - 9 sets of predictions:
     - C_α, C_β and side-chains centers
     - contact cutoffs: 6Å, 7Åand 8Å

I-TASSER

Pipeline

2. Structure assembly
   - Knowledge-based force field:
     1. general statistic terms from PDB (C_α/side chain correlations, H-bonds, hydrophobicity)
     2. spatial restraints from threading templates
     3. sequence-based contact prediction from SVMSEQ
   - 9 sets of predictions:
     - C_α, C_β and side-chains centers
     - contact cutoffs: 6Å, 7Åand 8Å

I-TASSER

Pipeline

2. Structure assembly
   - Knowledge-based force field:
     1. general statistic terms from PDB (C_α/side chain correlations, H-bonds, hydrophobicity)
     2. spatial restraints from threading templates
     3. sequence-based contact prediction from SVMSEQ
   - 9 sets of predictions:
     - C_α, C_β and side-chains centers
     - contact cutoffs: 6Å, 7Åand 8Å

I-TASSER

Pipeline

2. Structure assembly
   - Knowledge-based force field:
     1. general statistic terms from PDB (C_α/side chain correlations, H-bonds, hydrophobicity)
     2. spatial restraints from threading templates
     3. sequence-based contact prediction from SVMSEQ
   - 9 sets of predictions:
     - C_α, C_β and side-chains centers
     - contact cutoffs: 6Å, 7Åand 8Å

I-TASSER

Pipeline

3. Atomic model construction
   Full-atomic model built by REMO from C_α-traces
   - Basic backbone atoms: secondary structure specific backbone library
   - Refinement
     using: H-bonding
     clash/break-amendment
     I-TASSER restraints

     CHARMM22 potential

I-TASSER

Pipeline

3. Atomic model construction
   Full-atomic model built by REMO from C_α-traces
   - Basic backbone atoms: secondary structure specific backbone library
   - Refinement
     using: H-bonding
     clash/break-amendment
     I-TASSER restraints

     CHARMM22 potential

I-TASSER

Pipeline

3. Atomic model construction
   Full-atomic model built by REMO from C_α-traces
   - Basic backbone atoms: secondary structure specific backbone library
   - Refinement
     using: H-bonding
     clash/break-amendment
     I-TASSER restraints

     CHARMM22 potential

I-TASSER

Pipeline

4. Model selection
   - Ranking of reduced models according to cluster density
   - Selection of atomic models from REMO based on an empirical scoring function

I-TASSER

Pipeline

4. Model selection
   - Ranking of reduced models according to cluster density
   - Selection of atomic models from REMO based on an empirical scoring function

I-TASSER

Pipeline

4. Model selection
   - Ranking of reduced models according to cluster density
   - Selection of atomic models from REMO based on an empirical scoring function

I-TASSER

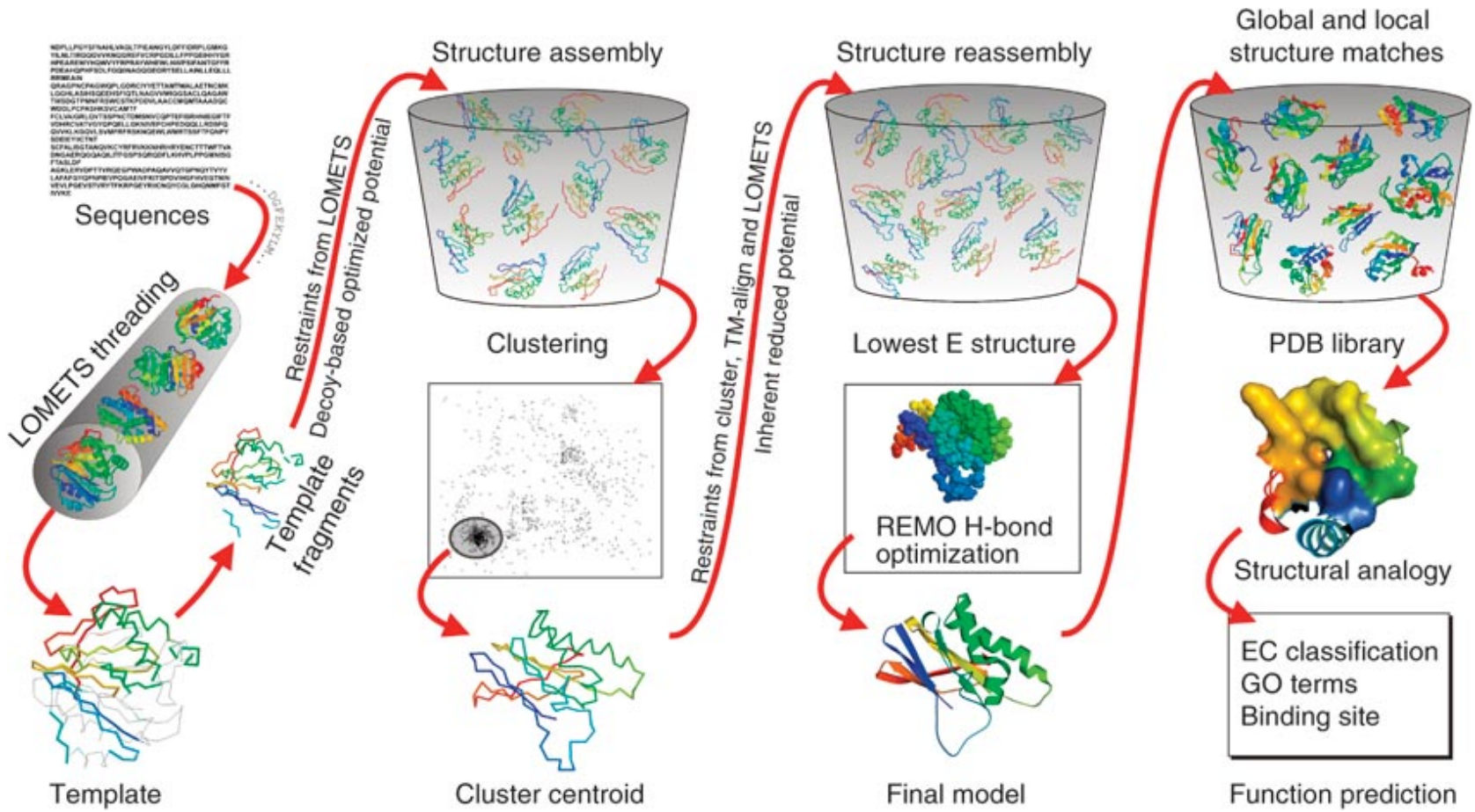
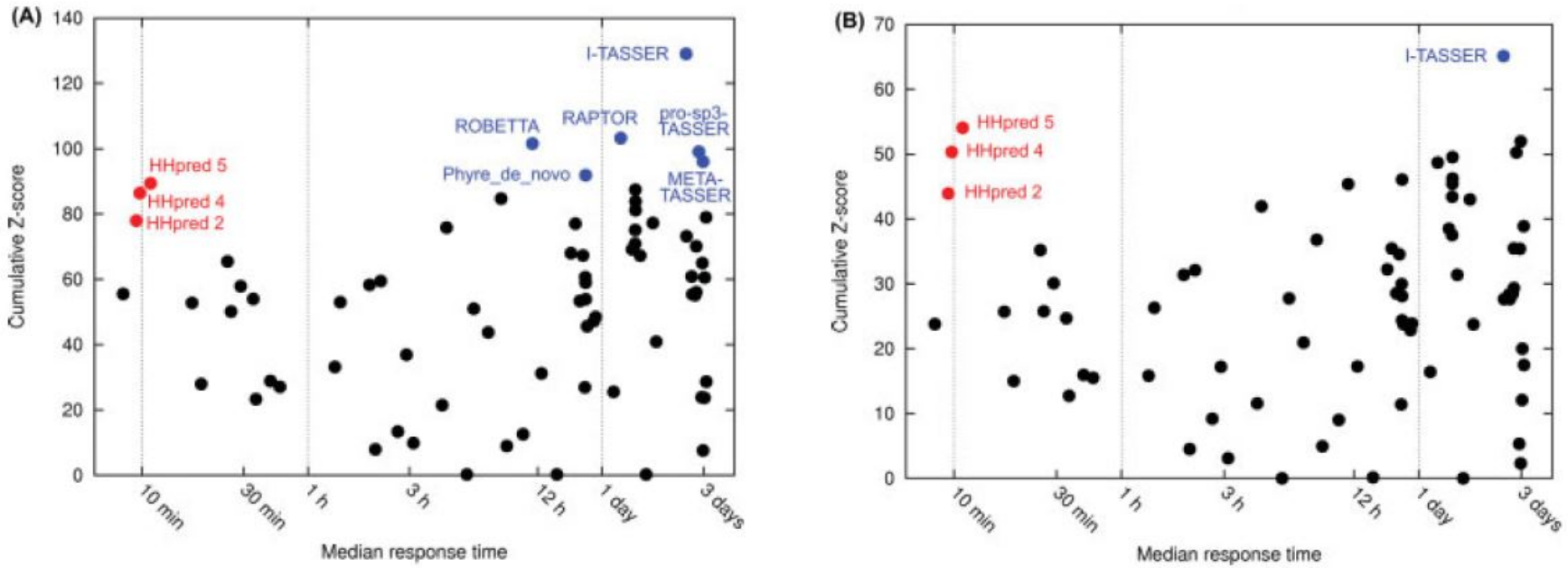


Figure: I-TASSER pipeline. (Roy, Kucukural, Yang 2010, I-TASSER: a unified platform for automated protein structure and function prediction.)

CASP8 results



(a) (b)

Figure: Model accuracy versus response time for the 72 servers in CASP8 (a) on the 164 target domains and (b) on the 85 single-domain targets. (Soeding et al. 2009, Fast and accurate automatic structure prediction with HHpred.)

Sources

- Elmar Krieger, Sander B. Nabuurs, and Gert Vriend (2003)
  Homology modeling
- Schwede T. (2003)
  SWISS-MODEL: An automated protein homology-modeling server.
- Schwede T.
  SWISS-MODEL server: http://swissmodel.expasy.org/
- Zhang Y. (2009)
  I-TASSER: fully automated protein structure prediction in CASP8
- Zhang Y.
  I-TASSER server: http://zhanglab.ccmb.med.umich.edu/I-TASSER/
- Soeding J. (2009)
  Fast and accurate automatic structure prediction with HHpred
- Soeding J.
  HHpred server: http://toolkit.tuebingen.mpg.de/hhpred