Atomic Contact Energies

Bioinformatics II

Adrian Geißler, Max Emil Schön

June 22, 2015

Tutor: Linus Backert

Structure

Theoretical Background

What does the energy function model?

Methods

Implementation of the function.

Results

Performance Evaluation

Discussion

Theoretical Background

Appetizer

from: http://thejobmouse.com

The Free Desolvation Energy

- Energy needed for transferring atoms from the solvent (water) into the protein's interior.
- One possible measure of protein stability
- The project:
  - Implement desolvation energy by (Zhang et al., 1997)
  - Evaluate on *CASP11* data

The Free Desolvation Energy

- Energy needed for transferring atoms from the solvent (water) into the protein's interior.
- One possible measure of protein stability
- The project:
  - Implement desolvation energy by (Zhang et al., 1997)
  - Evaluate on *CASP11* data

The Free Desolvation Energy

- Energy needed for transferring atoms from the solvent (water) into the protein's interior.
- One possible measure of protein stability
- The project:
  - Implement desolvation energy by (Zhang et al., 1997)
  - Evaluate on *CASP11* data

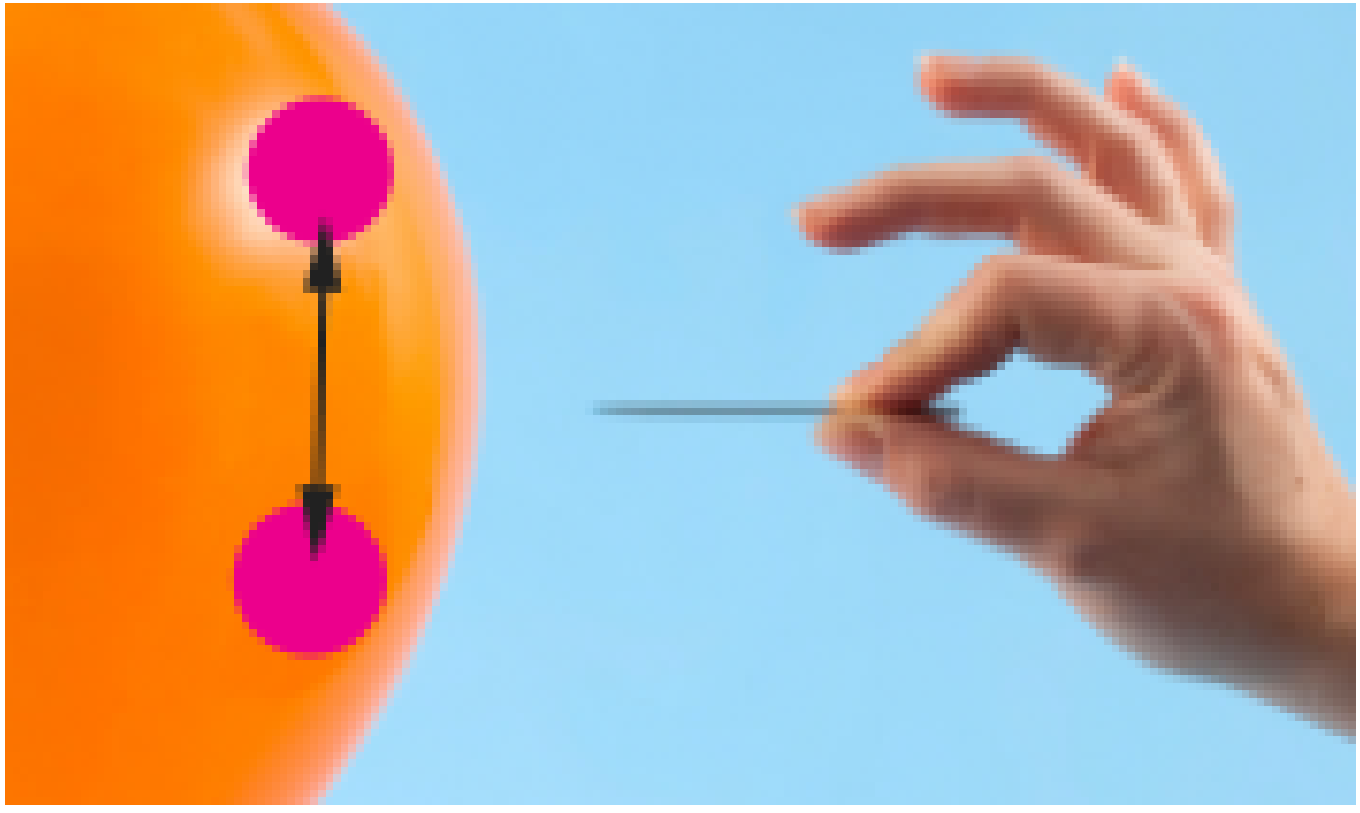
The Free Desolvation Energy

- Energy needed for transferring atoms from the solvent (water) into the protein's interior.
- One possible measure of protein stability
- The project:
  - Implement desolvation energy by (Zhang et al., 1997)
  - Evaluate on *CASP11* data

The Free Desolvation Energy

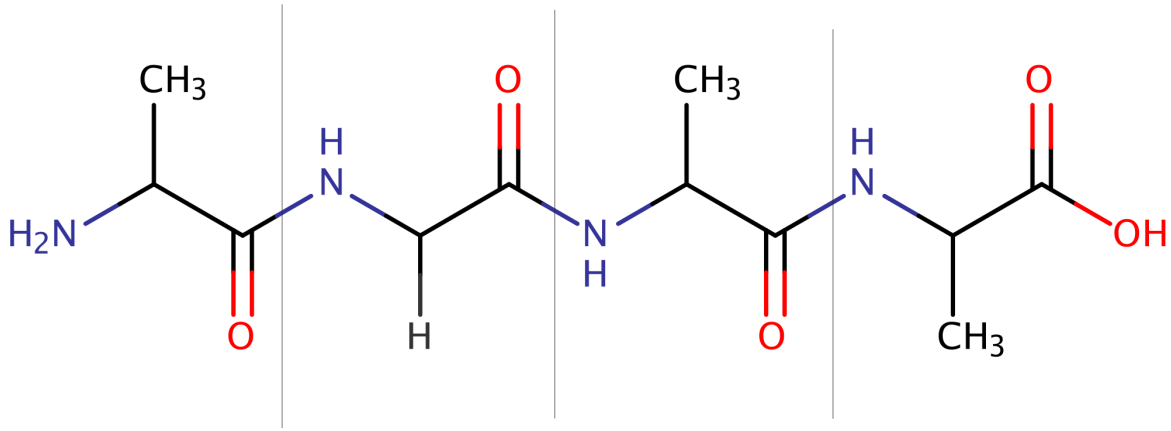
- Energy needed for transferring atoms from the solvent (water) into the protein's interior.
- One possible measure of protein stability
- The project:
  - Implement desolvation energy by (Zhang et al., 1997)
  - Evaluate on *CASP11* data

Atomic Contacts



Based on a picture from: `http://thejobmouse.com`

Atomic Contacts Pairs

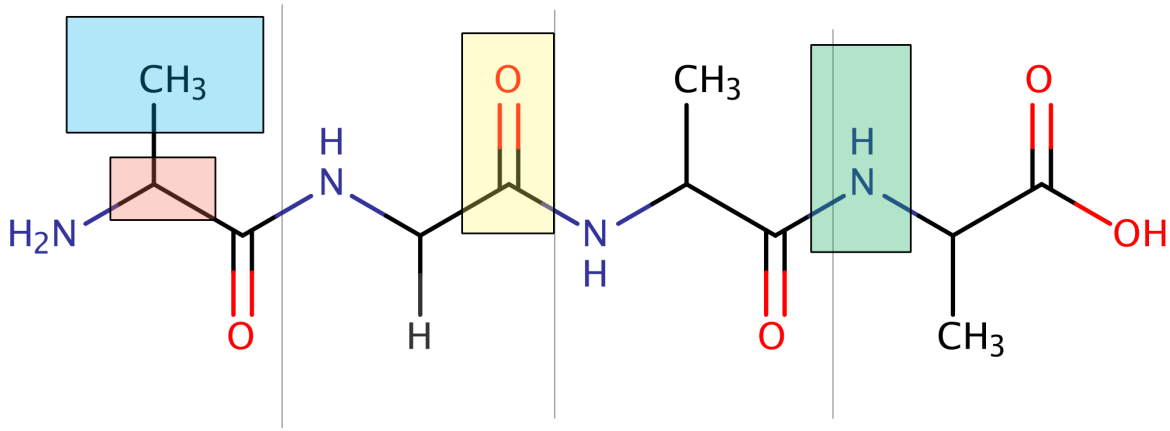


A valid contact pair:

- Only heavy atoms
- Distance below 6 Å
- More than 10 covalent bonds in between
  Estimated by connectivity class & residue index differences

Overall energy is a simple sum

Atomic Contacts Pairs

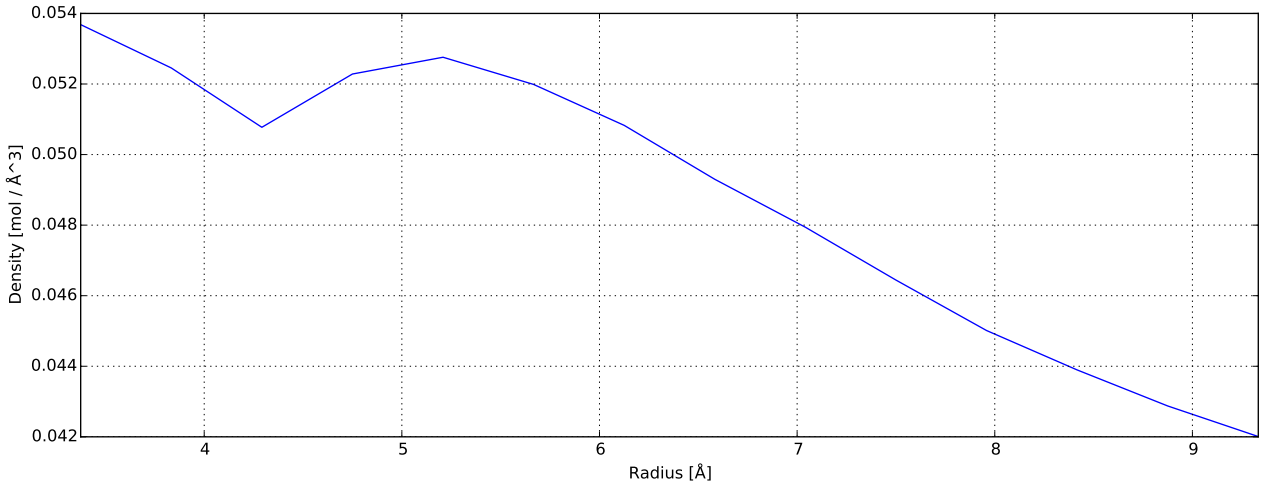


A valid contact pair:

- Only heavy atoms
- Distance below 6 Å
- More than 10 covalent bonds in between
  Estimated by connectivity class & residue index differences

Overall energy is a simple sum

Atomic Packaging



- Number density of interior atoms (SAS = 0)
- Relative to a sphere of variable radius
- Evaluated on non-homologous protein set

Methods

Implementation

- Implemented in Python
- PDB package from BioPython (Hamelryck and Manderick, 2003)
- Parallelized with multiprocessing package
- Datastructures from pandas
- Matrix for pairwise contacts from (Zhang et al., 1997)

Implementation

- Implemented in Python
- PDB package from BioPython (Hamelryck and Manderick, 2003)
- Parallelized with multiprocessing package
- Datastructures from pandas
- Matrix for pairwise contacts from (Zhang et al., 1997)

Implementation

- Implemented in Python
- PDB package from BioPython (Hamelryck and Manderick, 2003)
- Parallelized with multiprocessing package
- Datastructures from pandas
- Matrix for pairwise contacts from (Zhang et al., 1997)

Implementation

- Implemented in Python
- PDB package from BioPython (Hamelryck and Manderick, 2003)
- Parallelized with multiprocessing package
- Datastructures from pandas
- Matrix for pairwise contacts from (Zhang et al., 1997)

Implementation

- Implemented in Python
- PDB package from BioPython (Hamelryck and Manderick, 2003)
- Parallelized with multiprocessing package
- Datastructures from pandas
- Matrix for pairwise contacts from (Zhang et al., 1997)

Implementation cont'd

1. Load and pre-process Reference structure
2. Filter PDB file
3. For every possible contact pair, check connectivity criterion
4. sum over all pairwise energies

Implementation cont'd

1. Load and pre-process Reference structure
2. Filter PDB file
3. For every possible contact pair, check connectivity criterion
4. sum over all pairwise energies

Implementation cont'd

1. Load and pre-process Reference structure
2. Filter PDB file
3. For every possible contact pair, check connectivity criterion
4. sum over all pairwise energies

Implementation cont'd

1. Load and pre-process Reference structure
2. Filter PDB file
3. For every possible contact pair, check connectivity criterion
4. sum over all pairwise energies

RMSD

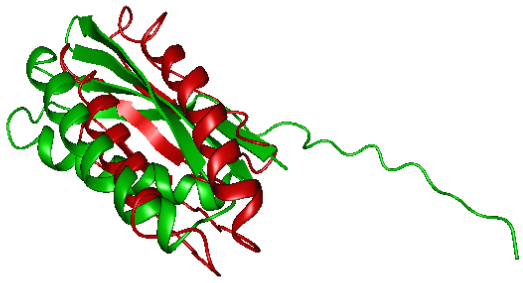
- Alignment of prediction and target structure based on superposition of $C\alpha$ atoms (again BioPython)
- RMSD as measure of deviation

RMSD

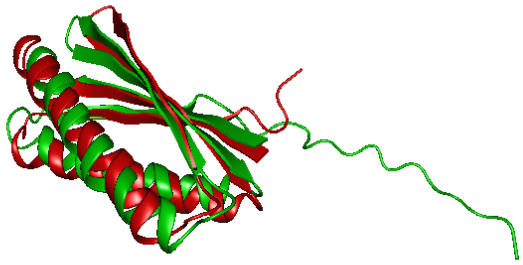
- Alignment of prediction and target structure based on superposition of $C\alpha$ atoms (again BioPython)
- RMSD as measure of deviation

Results

Comparison with Reference (1/2)



T0769-442



T0769-241

Comparison with Reference (2/2)

T0784-117

T0784-156

Energy v. RMSD

Table : For the dataset `T0784` and `T0769`, the tables show the computed contact energies and the RMSD counterparts sorted by energies (left) and by RMSD (right).

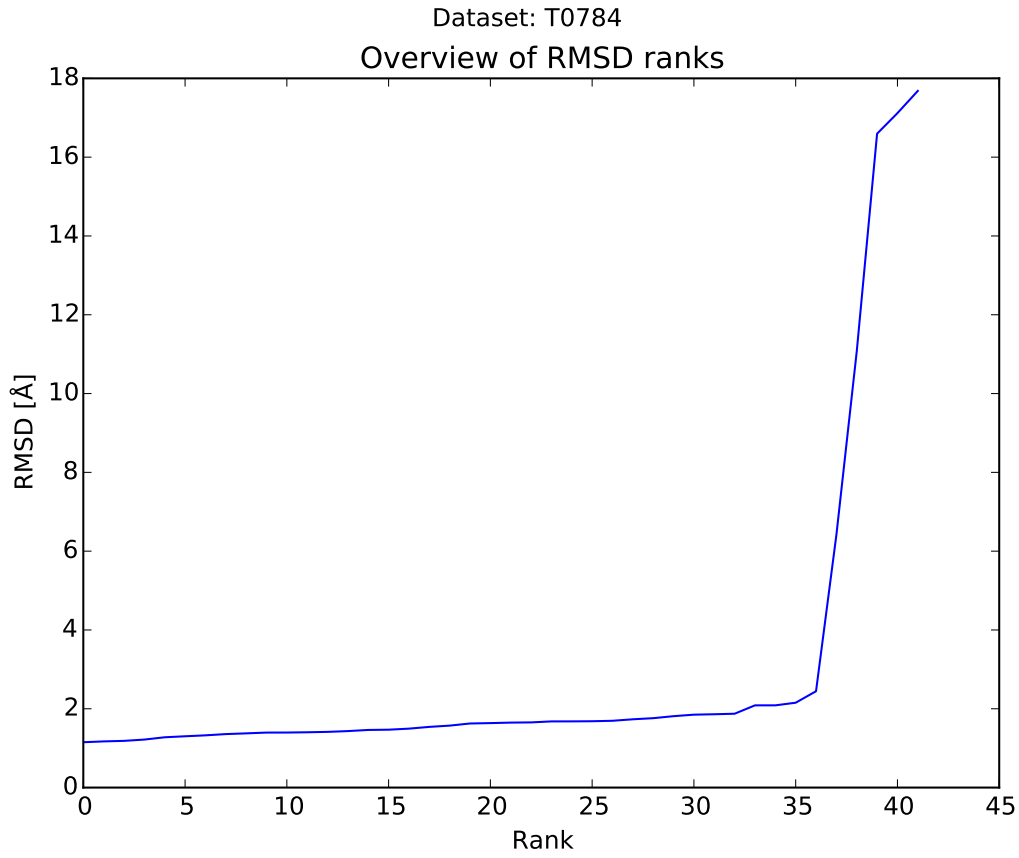
`T0784`	Energy in $\frac{kcal}{mol}$	RMSD
T0784TS156_1	−130.48	1.15
T0784TS420_1	−127.74	1.17
T0784TS499_1	−149.46	1.18
T0784TS237_1	−139.99	1.22
T0784TS268_1	−160.82	1.28

`T0784`	Energy in $\frac{kcal}{mol}$	RMSD
T0784TS117_1	−230.12	1.73
T0784TS008_1	−203.16	1.86
T0784TS251_1	−193.6	1.63
T0784TS038_1	−162.31	1.38
T0784TS268_1	−160.82	1.28

`T0769`	Energy in $\frac{kcal}{mol}$	RMSD
T0769TS241_1	−59.34	2.67
T0769TS368_1	−66.75	3.16
T0769TS258_1	−74.39	4.37
T0769TS361_1	−79.04	4.41
T0769TS186_1	−79.97	4.51

`T0769`	Energy in $\frac{kcal}{mol}$	RMSD
T0769TS442_1	−90.73	16.72
T0769TS155_1	−90.62	17.12
T0769TS044_1	−84.32	10.38
T0769TS169_1	−81.02	10.39
T0769TS317_1	−80.61	6.89

Investigation of correlation



Pearson: 0.19 – 0.44

Spearman: 0.15 – 0.33

Conclusion

Conclusion

- Trends of structural correspondence (except for `T0769`)
- Linear Correlation measures of energy and RMSD fail
  $\Rightarrow$ RMSD almost the same for first 2/3 of predictions
- RMSD measure not reliable
  $\Rightarrow$ CASP uses multiple measures
- Function by (Zhang et al., 1997) is fast! (only sums)
- But: 10 years old

Conclusion

- Trends of structural correspondence (except for `T0769`)
- Linear Correlation measures of energy and RMSD fail
  $\Rightarrow$ RMSD almost the same for first 2/3 of predictions
- RMSD measure not reliable
  $\Rightarrow$ CASP uses multiple measures
- Function by (Zhang et al., 1997) is fast! (only sums)
- But: 10 years old

Conclusion

- Trends of structural correspondence (except for `T0769`)
- Linear Correlation measures of energy and RMSD fail
  $\Rightarrow$ RMSD almost the same for first 2/3 of predictions
- RMSD measure not reliable
  $\Rightarrow$ CASP uses multiple measures
- Function by (Zhang et al., 1997) is fast! (only sums)
- But: 10 years old

Conclusion

- Trends of structural correspondence (except for `T0769`)
- Linear Correlation measures of energy and RMSD fail
  $\Rightarrow$ RMSD almost the same for first 2/3 of predictions
- RMSD measure not reliable
  $\Rightarrow$ CASP uses multiple measures
- Function by (Zhang et al., 1997) is fast! (only sums)
- But: 10 years old

Conclusion

- Trends of structural correspondence (except for `T0769`)
- Linear Correlation measures of energy and RMSD fail
  $\Rightarrow$ RMSD almost the same for first 2/3 of predictions
- RMSD measure not reliable
  $\Rightarrow$ CASP uses multiple measures
- Function by (Zhang et al., 1997) is fast! (only sums)
- But: 10 years old

Thank you for your attention!

Bibliography

Hamelryck T, Manderick B (2003)

PDB file parser and structure class implemented in Python.

Bioinformatics 19:2308–2310.

Zhang C, Vasmatzis G, Cornette JL, DeLisi C (1997)

Determination of atomic desolvation energies from the structures of crystallized proteins.

Journal of molecular biology 267:707–726.