Atomic Contact Energies

Max Emil Schn, Adrian Geiler

June 17, 2015

1 Introduction

Energy functions are used to model protein 3D structures. Aspects that influence the free energy of proteins are van-der-Waals forces, electrostatics, dipole interactions, and torsions. Zhang et al. (1997) described a method to approximate the solvation energy based on atom contacts, rather than residue interactions, as proposed by Miyazawa and Jernigan (1996).

Their approach is to assign one of 18 atom types to each heavy atom. Based on these types, the energetic contribution of each contact pair to the overall energy is derived. A contact pair is defined by Zhang et al. (1997) as two atoms that have a distance below 6 Å and that are at least ten covalent bonds away from each other. The covalent distance is approximated with respect to connectivity classes without considering bonds in side-chains.

Here, we implement this algorithm and assess its quality on various protein prediction data sets.

2 Material and Methods

We implemented the algorithm in the Python language. In addition, we used Biopython's pdb processing functionality (Hamelryck and Manderick, 2003).

The pairwise contact energies were adopted from Zhang et al. (1997). We used the author's cutoff distance of 6 Å, because we were able to verify this value based on the atomic packaging of a large set of non-homologous experimental protein structures (see Figure 3)

The quality of the implemented function was assessed on a subset of the submissions to the eleventh CASP competition (Critical Assessment of protein Structure Prediction) [1]. The datasets in question were T0762, T0769, T0776, T0784. We evaluated each prediction by superimposing its backbone Cα atoms onto the reference and computing the RMSD, which we

[1] `http://www.predictioncenter.org/casp11` (last accessed: June 17, 2015)

Table 1: 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

compared to the RMSD values listed on the CASP website [2]. Afterwards, we investigated correlations between the RMSD and the computed energies. The python package SciPy (Oliphant, 2007) was used for the computation of the Pearson and the Spearman correlation coefficients.

3 Results

For every target structure we visualized the prediction with the lowest RMSD and the one with the lowest free contact energy in BALLView (Moll et al., 2006). Example visualizations of datasets T0679 and T0784 are shown in Fig. 1

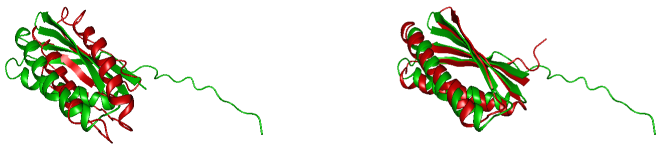
The relationship between ranked contact energy prediction and the RMSD are shown in Fig. 2.

Pearson's correlation coefficients between our free contact energies and the energies calculated by the CASP experiment ranged from 0.19 to 0.44. When comparing our energies to the corresponding ranks from CASP, the Spearman's correlation coefficients ranged from 0.15 to 0.33 (Table 1). We also manually compared our RSMD calculations with those from CASP and found that they largely correspond well to one another.

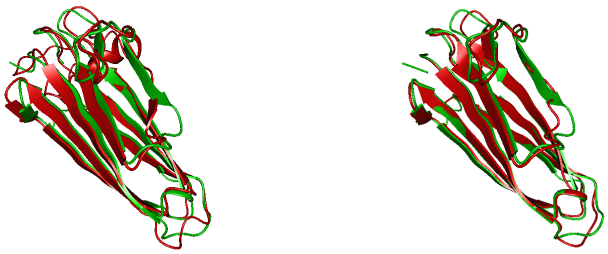
4 Discussion

After implementing the algorithm of Zhang et al. (1997), we were able to reproduce the threshold value of 6 Å on a non-homologous structure set. A relative increase of density for radii around this value was observable (see Figure 3) which could be due to the relevance for atomic contacts.

[2] `TODO` (last accessed: June 17, 2015)



(a) T0769 442 (b) T0769 241



(c) T0784 117 (d) T0784 156

Figure 1: Best scoring predictions of T0769 and T0784 evaluated using (a,c) atomic contact energies and (b,d) the RMSD between prediction and experimental structure. green: target experimental stucture. red: predicted structure.

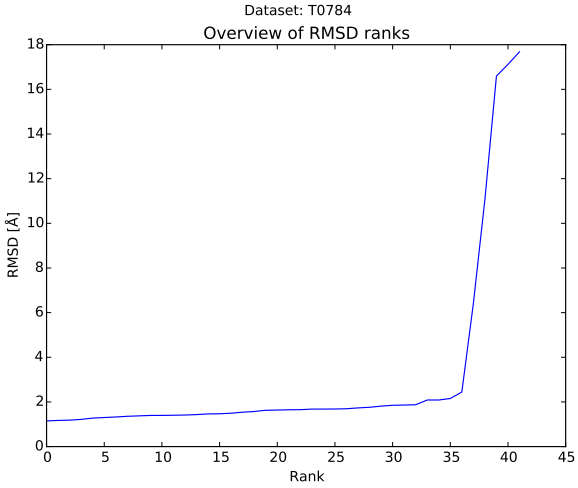


Figure 2: Calculated RMSD values are plotted against the rank of the corresponding prediction in the original CASP11 experiment `T0784`.

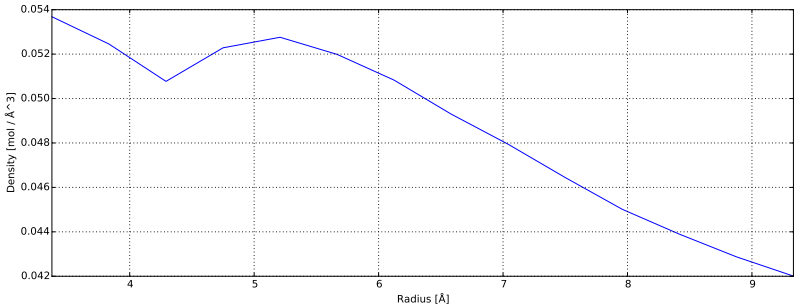


Figure 3: Shown is the atomic packaging estimated by the number density according to a radius around an interior atom. An atom is considered to be buried if the accessible surface area (SAS) is zero.

The correlation metrics Pearson and Spearman did not indicate a linear correlation between RMSD and the computed energies. However, visual inspections of the RMSD distribution suggest that these metrics are bound to fail, because significant changes were only observed for the lower ranking third of the predicted structures. The original CASP11 evaluations take this property of RMSD into account, as they utilize a multitude of other quality measures.

Although the correlation coefficients are low, plots of the superimposed structures show trends of consensus with the reference structure. Therefore, we suggest that the energy function of Zhang et al. (1997) is suitable to quickly assess the quality of folds, even though the resulting values will not be able to compete with more current free energy functions.

References

T. Hamelryck and B. Manderick. PDB file parser and structure class implemented in Python. *Bioinformatics*, 19(17):2308–2310, 2003.

S. Miyazawa and R. L. Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *Journal of molecular biology*, 256(3):623–644, 1996.

A. Moll, A. Hildebrandt, H.-P. Lenhof, and O. Kohlbacher. BALLView: a tool for research and education in molecular modeling. *Bioinformatics*, 22(3):365–366, 2006.

T. E. Oliphant. SciPy: Open source scientific tools for Python. *Computing in Science and Engineering*, 9:10–20, 2007.

C. Zhang, G. Vasmatzis, J. L. Cornette, and C. DeLisi. Determination of atomic desolvation energies from the structures of crystallized proteins. *Journal of molecular biology*, 267(3):707–726, 1997.