Contacts defined by interaction energy extract essential thermodynamics from structures of globular proteins

Boris Fačkovec and Jiří Vondrášek

Introduction

Residue-residue contacts have been successfully used in lattice models and lot of effort has been put to structure representation in terms of contact matrices. The approach of Miyazawa and Jernigan [MJ] assigns knowledge-based free energy to a geometrical contact. Here we present an alternative approach. We determine, whether there is a contact (significantly strong interaction, measured by force field interaction energy) between two protein fragments of certain type.

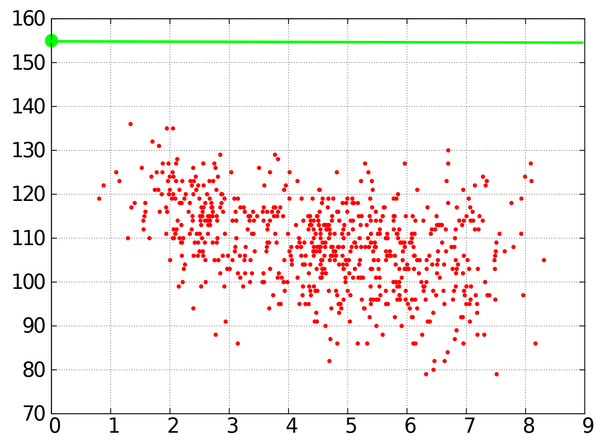
In previous work [FV] we have found that distributions of interaction energies between certain types of protein fragments have structure that allows to distinguish strong, "productive", contacts from interactions between neighbours which do not contribute to stability ("bulk" interactions). The number of contacts in a protein correlates with thermostability for a given class of proteins for given pH.

Here we show that the number of contacts in a structure can be an efficient scoring function for protein structure prediction.

Number of contacts as a scoring function for structure prediction

The number of contacts was tested on Decoys 'R' Us [SL] decoy set. Performance is very good for all decoy sets except for hg_structal (18/29 native structures scored best) and ig_structal (7/61 native structures scored best).

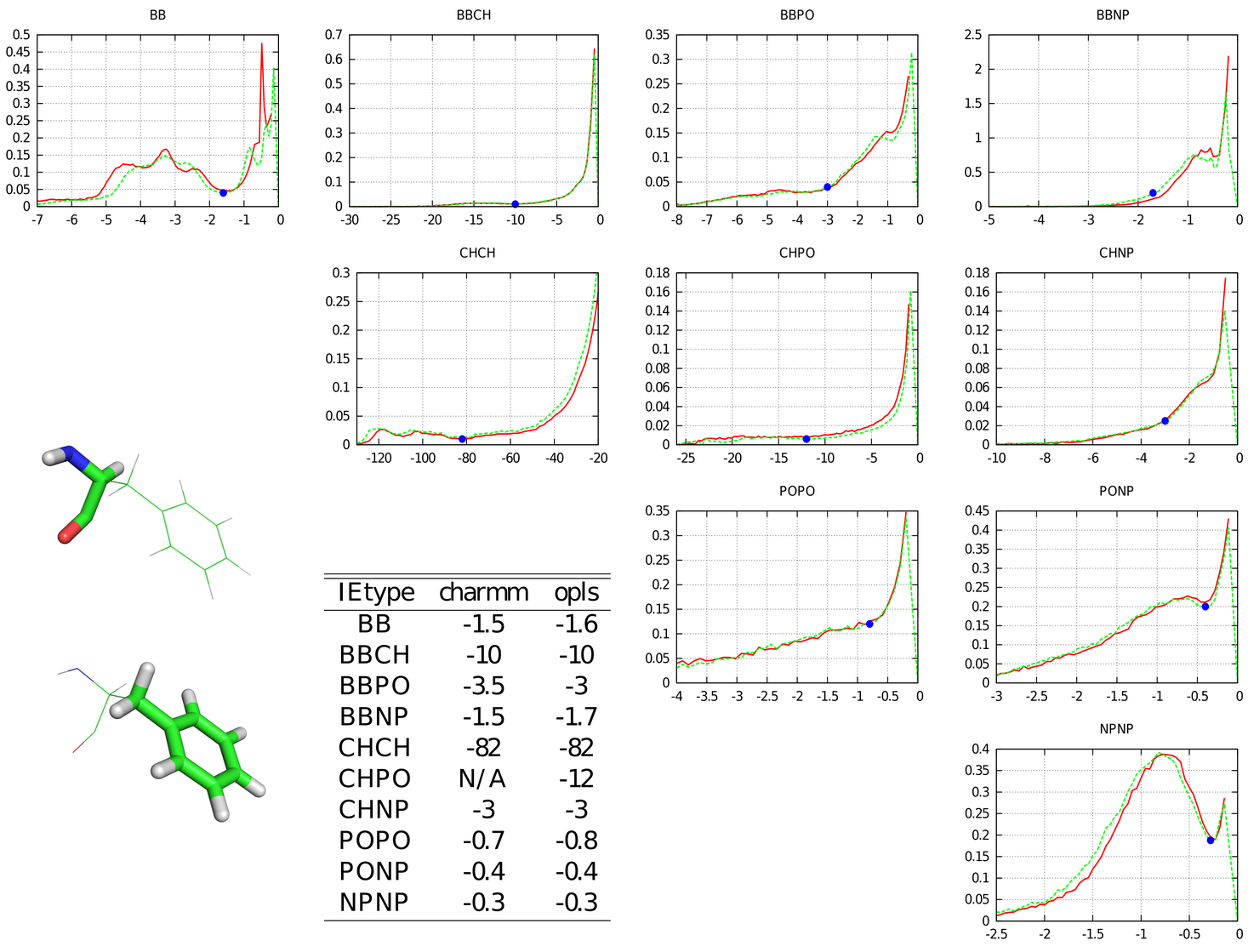
prot	OPLS rank	OPLS z-score	CHARMM rank	CHARMM z-score	TE-13 rank	TE-13 z-score	Rosetta rank	DOPE rank
				4state_reduced				
1ctf	1	5.18	1	4.65	1	4.20	1	1
1r69	1	5.44	1	4.86	1	4.06	2	1
1sn3	3	2.48	2	3.19	6	2.70	1	1
2cro	1	5.34	1	4.57	1	3.48	5	1
3icb	1	2.85	2	2.83			6	1
4pti	1	3.88	1	3.99	7	2.43	1	1
4rxn	1	3.63	1	3.77	16	1.79	1	1
				fisa				
1fc2	404	-0.83	451	-1.39	16	1.67	158	375
1hdd-c	3	2.39	2	2.72	1	4.35	90	1
2cro	1	5.31	1	5.12	1	4.00	26	1
4icb	1	6.13	1	6.12			1	1
				fisa_casp3				
1bg8-a	1	5.17	1	4.76	3	2.98	1068	1
1bl0	1	5.19	1	5.07	3	2.80	960	1
1eh2	26	2.32	136	1.47				
1jwe	1	6.53	1	6.28	1	6.04	1177	1
				lattice_ssfit				
1beo	1	8.86	1	11.75			1	1
1ctf	1	9.80	1	8.88	1	6.17	1	1
1dkt-a	1	6.32	1	6.46	2	3.92	1	1
1fca	1	3.88	1	4.57	36	2.25	1	1
1nkl	1	7.18	1	7.06	1	4.51	1	1
1pgb	1	13.51	1	12.62	1	4.13	1	1
1trl-a	1	7.20	1	6.77	1	3.63	45	1
4icb	1	6.41	1	6.91			1	1



UP: number of contacts as a function of RMSD (in angstroms) from native structure 1CRO (green), which is clearly separated.
LEFT: performance of number of contacts compared to two of the successful score functions [TE] [SS]

Definition of contact

- Only non-bonded intra-molecular interactions were considered.
- Protein was fragmented into 2 N - n(Gly) fragments.
- Fragments are classified according to the range of their interactions into charged (DEHKR, denoted CH), polar (NQSTWY, PO), non-polar (ACFILMPV, NP) and backbone fragments (BB).
- After a steepest descent optimisation, interaction energies (IE) were calculated between fragments (no capping) using OPLS or CHARMM force field.
- Significantly strong interactions were found in IE distributions were found in distributions for representative structures from PDB.



IEtype	charmm	opls
BB	-1.5	-1.6
BBCH	-10	-10
BBPO	-3.5	-3
BBNP	-1.5	-1.7
CHCH	-82	-82
CHPO	N/A	-12
CHNP	-3	-3
POPO	-0.7	-0.8
PONP	-0.4	-0.4
NPNP	-0.3	-0.3

Discussion

High performance of the number of productive contacts as a scoring function can be justified by a (handwaving) **explanation**:
- Solvation energies for the tested decoys with the same primary structure is similar to the solvation energy of the corresponding native state.
- Tested decoys have little conformational clashes, so absence of their contribution in the scoring function does not cause errors.
- Strong contacts withstand vibrations of the native state and are representative interactions holding the structure together.
Low sensitivity to clashes makes the method robust to small conformational changes.

Implications:
- Evidence that contacts defined by interaction energy extract useful thermodynamic information about the structure and can then be used for further bioinformatic studies (SCOP/CATH classification, structure comparison).
- We can assess the contribution of each type of interaction to determination of the native state.
- We speculate that a protein can be viewed as a chain zipped by contacts.

Future work:
- Make the scoring function continuous - scaling IE's by the contact threshold energy.
- Decrease computational cost by setting distance thresholds for IE calculations.
- Develop an analogous method for identification of strong protein-protein contacts (see also [KV]).
- Compare the contact matrix with elastic networks and evolutionary conserved contacts.
- Enrich physical structure prediction methods [MF]

Ackowledgement

Bakala Foundation
King's College, Cambridge
Grant No. P208/10/0725
(Czech Science Foundation)
Grant No. LH11020 (MSMT)
subv. dev. res. org.
RVO: 61388963.

References

[FV] Fačkovec & Vondrášek, 2012, JPCB 116, 12651
[KV] Kysilka & Vondrášek, 2012, JMolRecog 25, 604
[MF] Moult et al., 2014, Proteins 82, 1
[MJ] Miyazawa & Jernigan, 1996, JMolBiol 256, 623
[SL] Samudrala & Levitt, 2000, ProtSci 9, 1399
[SS] Shen & Šali, 2006, ProtSci 15, 2507
[TE] Tobi & Elber, 2000, Proteins 41, 40

Contact

BF: bf269@cam.ac.uk, Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, United Kingdom
JV: jiri.vondrasek@uochb.cas.cz, Institute of Organic Chemistry and Biochemistry, Flemingovo nam 2, 160 00 Praha 6, Czech Republic