



Bridging Biophysics and Bioinformatics General Trends of Intramolecular Interactions in Globular Proteins

Nové Hradý, 4th May 2011

Boris Fačkovec
Institute of Organic Chemistry and Biochemistry,
AS CR
borislavujo@gmail.com

- Our approach to protein analysis

- Entropy of system

- **Enthalpy of system**

- Denatured state

- **Native state**

- Solvent-solvent interactions

- Solvent-protein interactions

- **Protein intramolecular interactions**

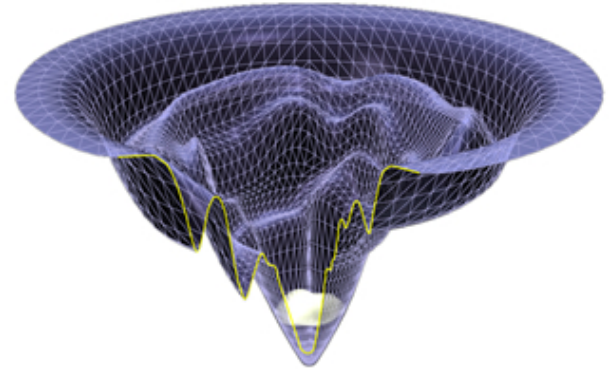
- Covalent interactions

- **Non-covalent interactions**

- We calculate interaction energies of non-covalent interactions

- We characterize selected data from structural database by physical means

$$\Delta G = \Delta H - T \Delta S$$



- Protein containing N amino acids is split into 2N groups of atoms (each residue – 2 groups)
- 4 types of atom groups (**fragments**) based on their **physical character**

- Backbone (BB)

- Dipoles with special organization

- Side chain - charged (CH - D,E,K,R,H)

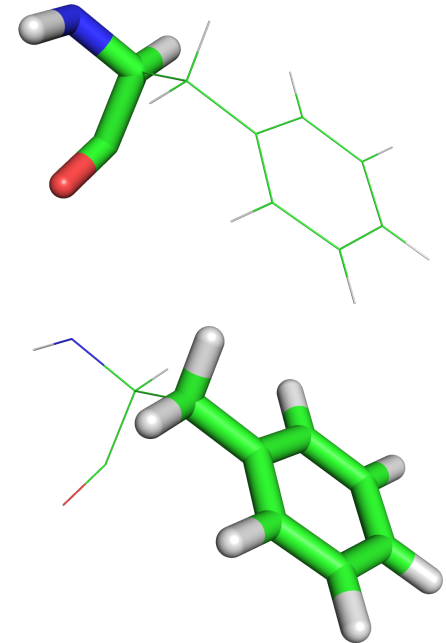
- ions

- Side chain - polar (PO - N,Q,T,S)

- dipoles

- Side chain - non-polar (NP - G,A,L,I,V,C,M,P,F,Y,W)

- quadrupoles
 - van der Waals



→ 10 types of interactions

- We acquire **interaction energy** (IE) for each pair of fragments
- **OPLS Force field** – interaction energy between 2 fragments is a sum of their atomic pairwise contributions

- Coulombic terms
 - Gas phase ($\epsilon_r=1$)

$$E_{coulomb} = k \frac{q_1 q_2}{r}$$

- Lennard-Jones terms
 - Subsequent backbone interactions set zero

$$E_{LJ} = \epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

- Structure set – 1358 single chain proteins, no ligands, 70% sequence identity removed

IEM and RIE construction – backbone-backbone interactions

- IE gathered in **interaction energy matrix (IEM)**
- Protein of N residues → NxN matrix of pairwise interactions
- We have 10 types of interactions → 10 IEMs for each protein

BB IEM	ALA 1	ALA 2	GLN 3	SER 4	VAL 5	ASP 6	GLN 7	LEU 8	ILE 9	LYS 10
ALA 1	0.00	0.00	-2.90	0.56	0.12	0.81	-0.82	0.74	0.47	0.43
ALA 2	0.00	0.00	0.00	-0.34	0.02	-0.11	-0.15	-0.01	-0.03	-0.05
GLN 3	-2.90	0.00	0.00	0.00	-1.09	-0.02	0.39	0.21	0.00	0.09
SER 4	0.56	-0.34	0.00	0.00	0.00	1.26	-0.95	-3.11	-0.06	-0.21
VAL 5	1.12	0.02	-1.09	0.00	0.00	0.00	0.01	-1.09	-3.79	-0.28
ASP 6	0.81	-0.11	-0.02	1.26	0.00	0.00	0.00	-0.05	-1.55	-3.65
GLN 7	0.82	-0.15	0.39	-0.95	0.01	0.00	0.00	0.00	-0.15	-1.43
LEU 8	0.74	-0.01	0.21	-3.11	-1.09	-0.05	0.00	0.00	0.00	-0.06
ILE 9	0.47	-0.03	0.00	-0.06	-3.79	-1.55	-0.15	0.00	0.00	0.00
LYS 10	0.43	-0.05	0.09	-0.21	-0.28	-3.65	-1.43	-0.06	0.00	0.00

- Residue interaction energy (RIE)
- Energy content of one residue
- Sum of values in one line of IEM

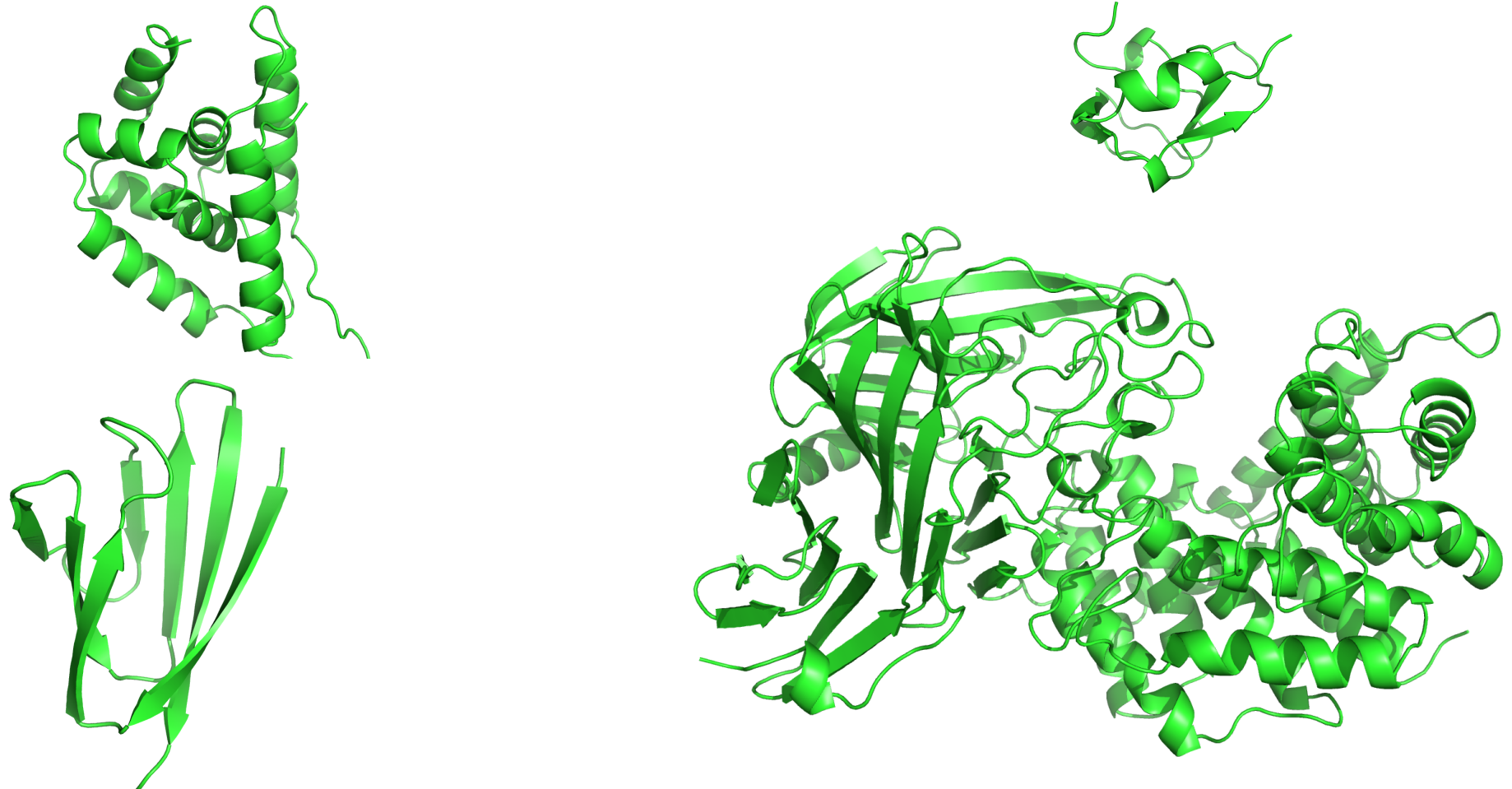
IEM and RIE construction – nonpolar-nonpolar interactions

NPNP IEM	ALA 1	ALA 2	GLN 3	SER 4	VAL 5	ASP 6	GLN 7	LEU 8	ILE 9	LYS 10
ALA 1	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ALA 2	0.01	0.00	0.00	0.00	-0.12	0.00	0.00	0.00	-0.22	0.00
GLN 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SER 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VAL 5	0.00	-0.12	0.00	0.00	0.00	0.00	0.00	-0.54	-0.32	0.00
ASP 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GLN 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LEU 8	0.00	0.00	0.00	0.00	-0.54	0.00	0.00	0.00	0.02	0.00
ILE 9	0.00	-0.22	0.00	0.00	-0.32	0.00	0.00	0.02	0.00	0.00
LYS 10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

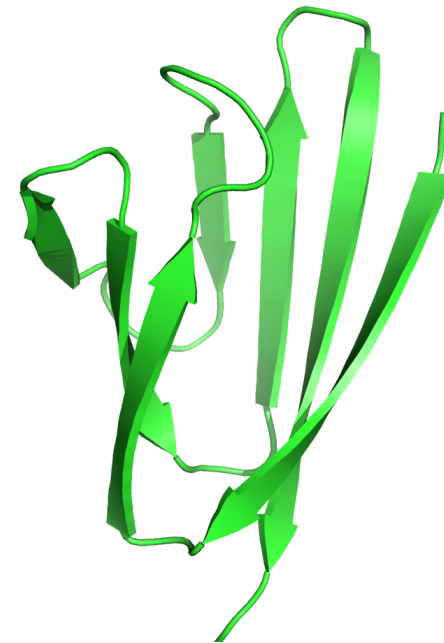
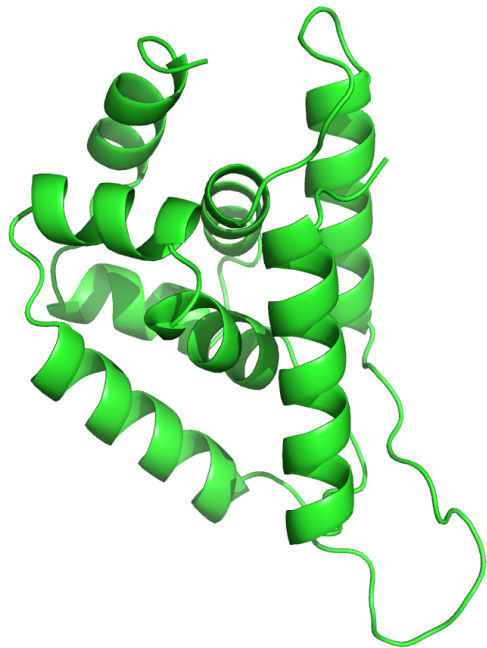


	NPNP RIE
ALA 1	0.02
ALA 2	-0.33
GLN 3	0.00
SER 4	0.00
VAL 5	-0.98
ASP 6	0.00
GLN 7	0.00
LEU 8	-0.52
ILE 9	-0.52
LYS 10	0.00

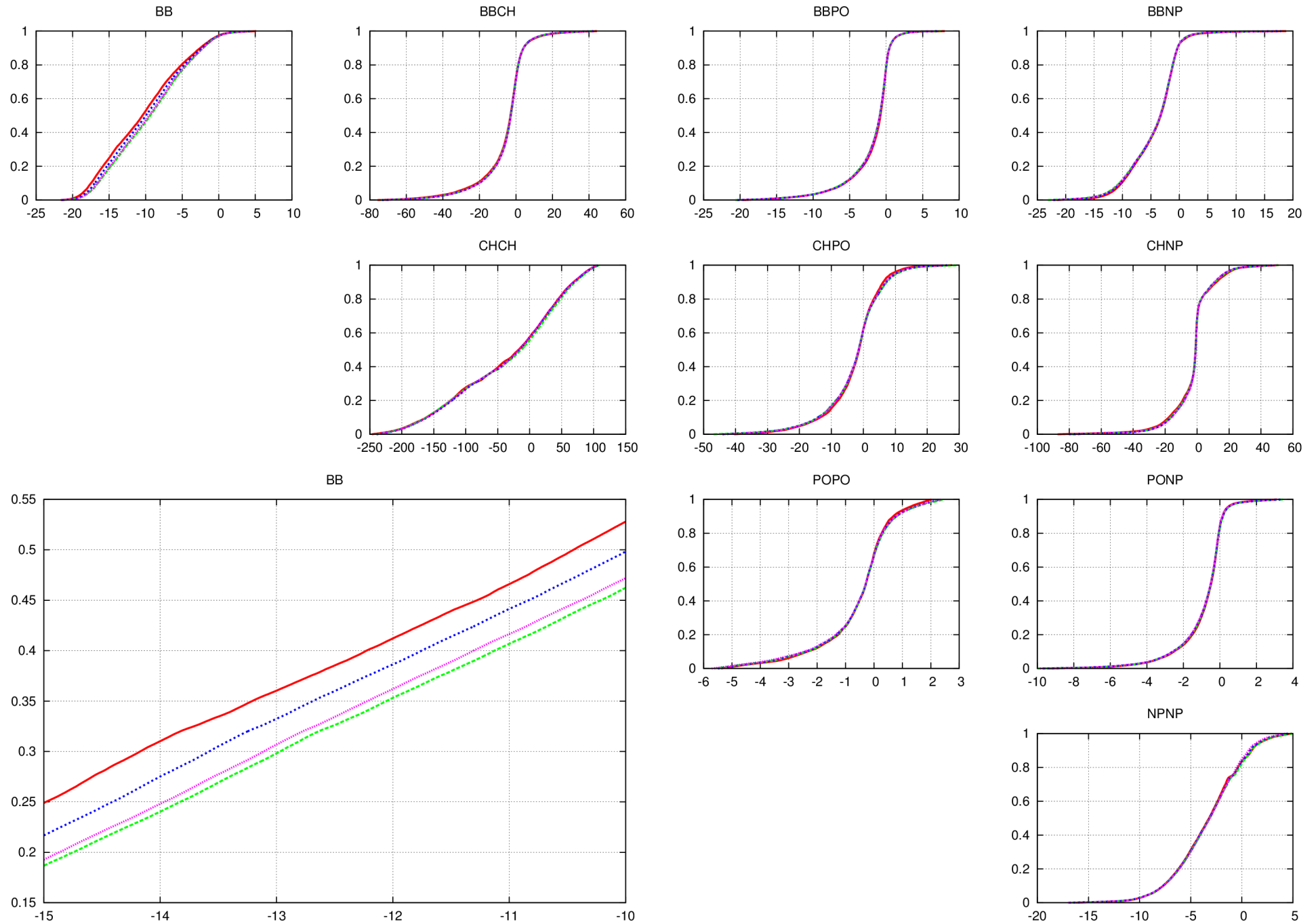
- Our task – statistics of RIEs calculated for large set of structures from database = **distributions**
- Dependence of RIE distributions on
 - **secondary structure content**
 - **size**



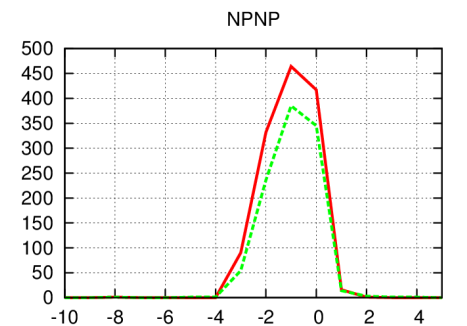
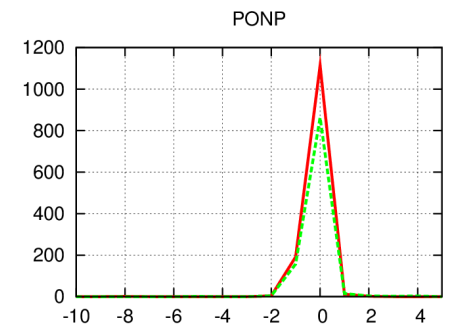
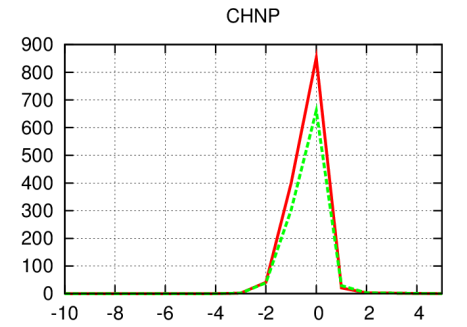
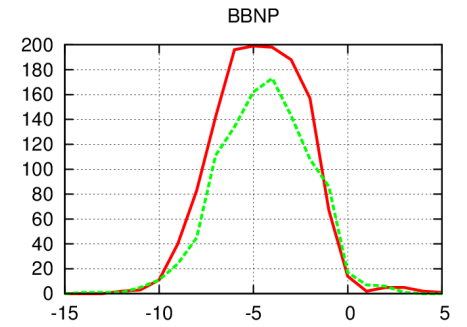
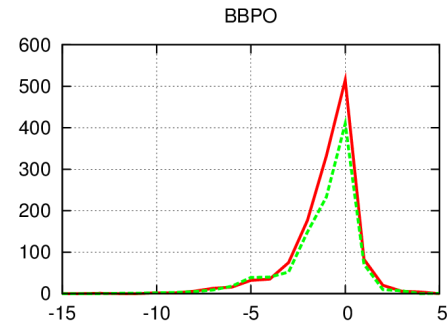
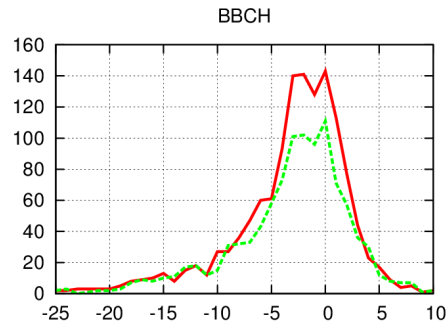
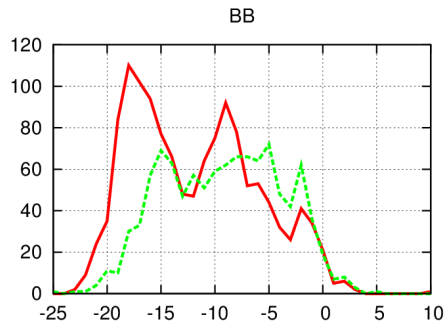
Effect of **secondary structure** content on RIE distributions



Average RIE distributions in proteins. Cumulative distribution function in proteins against RIE / [kcal]/mol. Red – CATH a, blue – SCOP a, green – CATH b, magenta – SCOP b.
In the left bottom corner – zoom on BB RIE distributions.



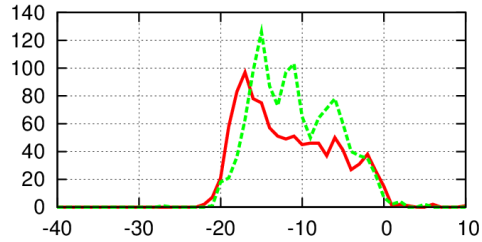
RIE distributions of alanine sampled through whole structure set. ALA contributes to 7 types of interactions.



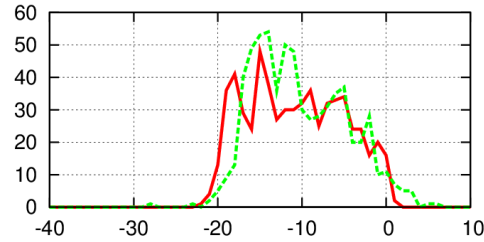
BB RIE distributions for each amino acid.

Red line - CATH mainly a proteins, green line – CATH mainly b

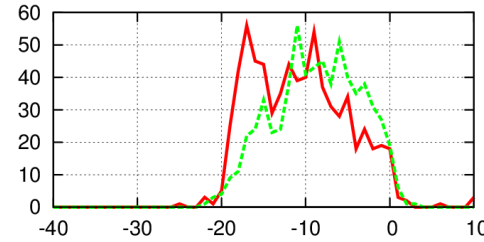
VAL



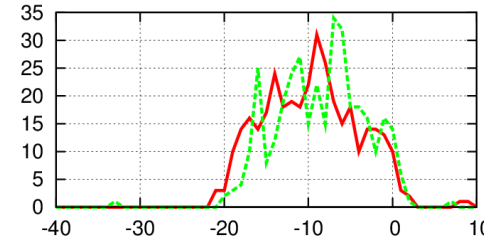
TYR



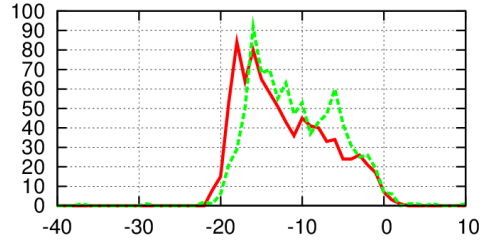
GLN



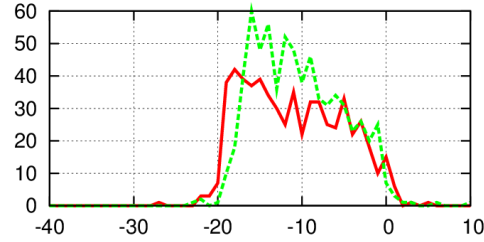
HIS



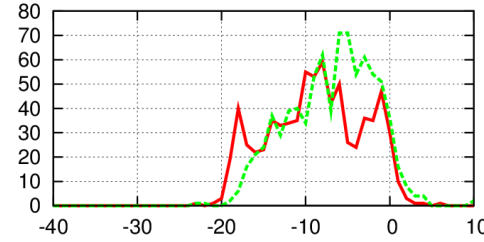
ILE



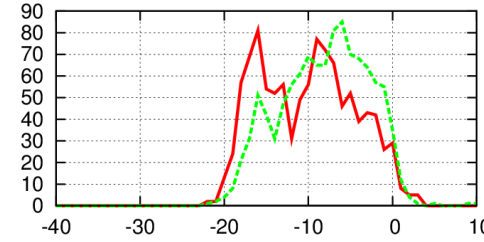
PHE



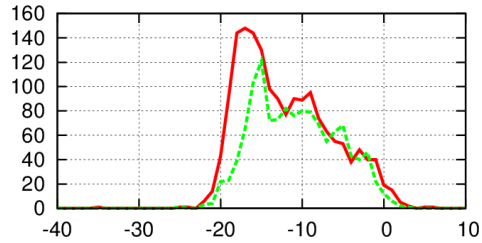
ASN



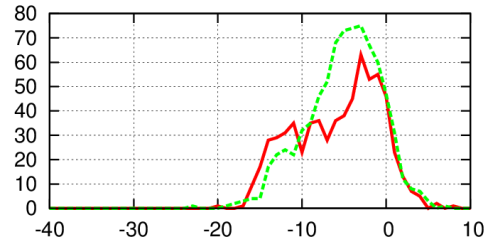
LYS



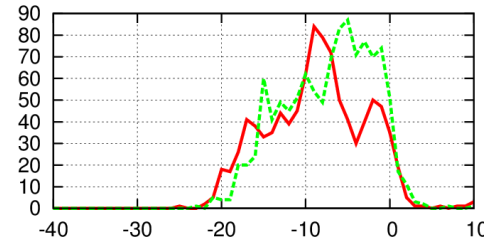
LEU



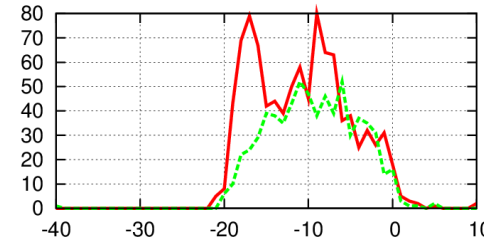
PRO



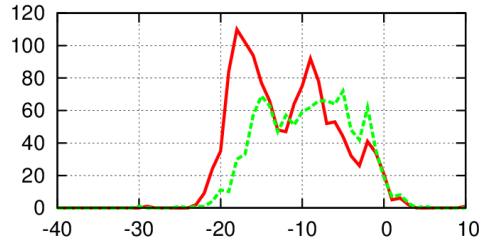
SER



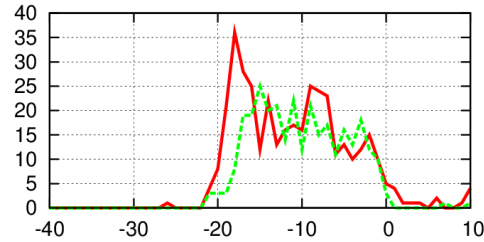
ARG



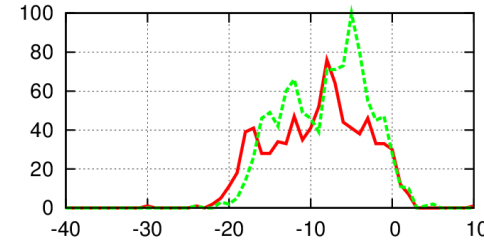
ALA



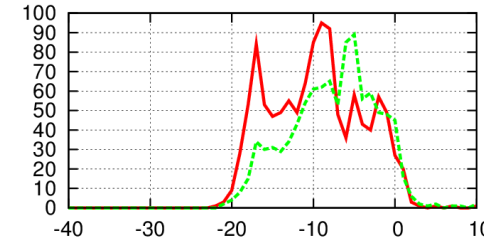
MET



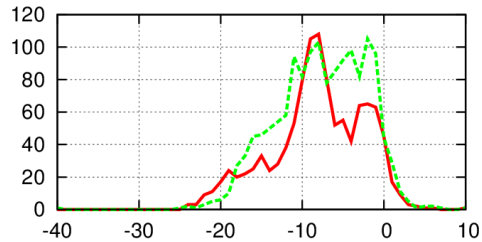
THR



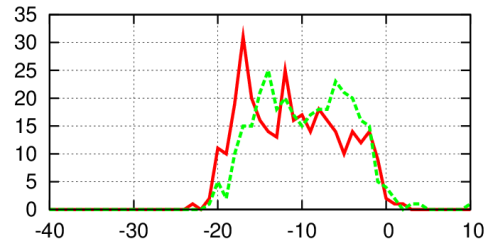
GLU



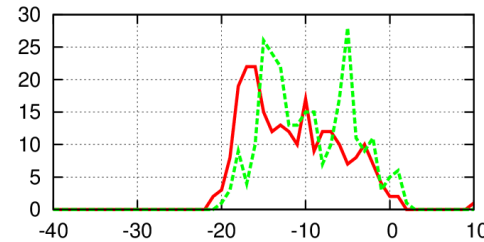
GLY



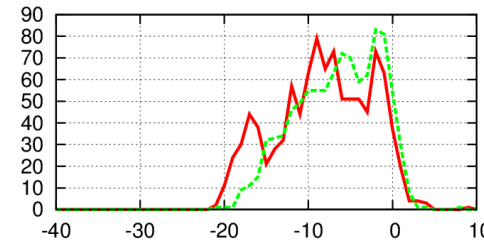
CYS



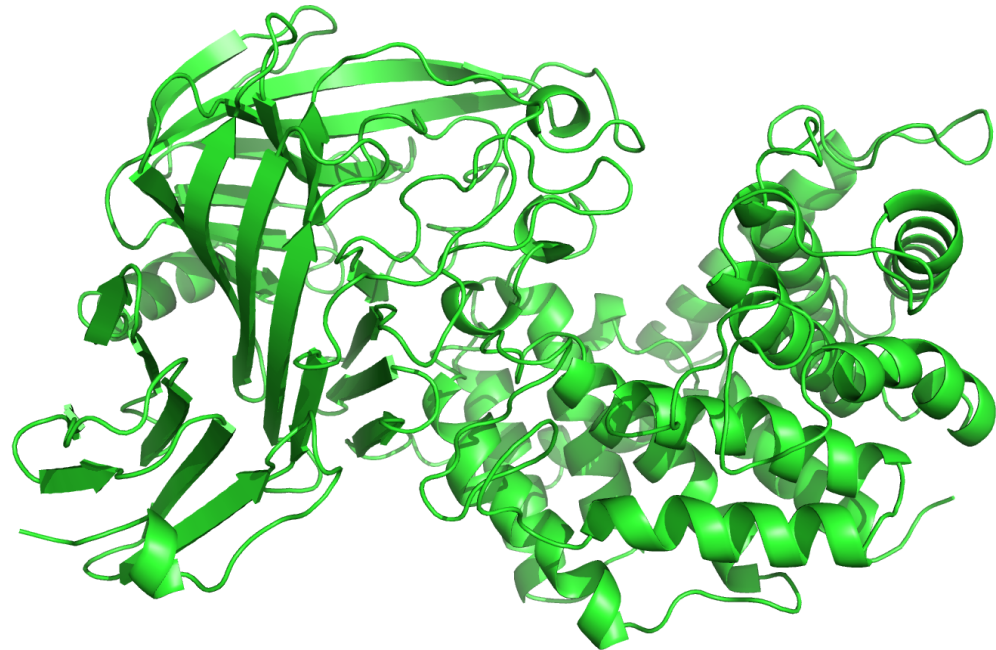
TRP



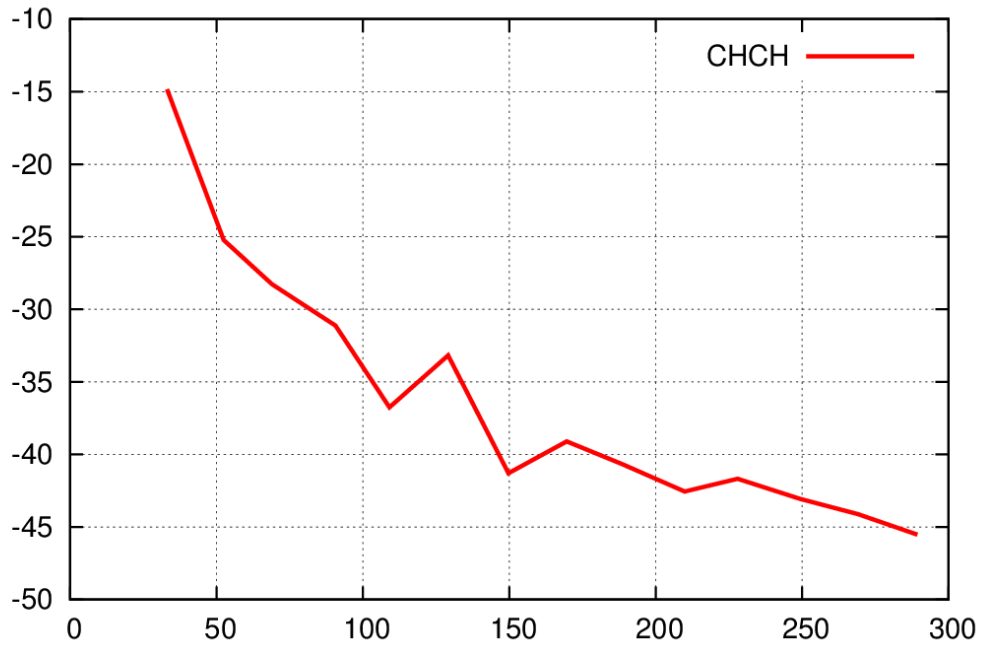
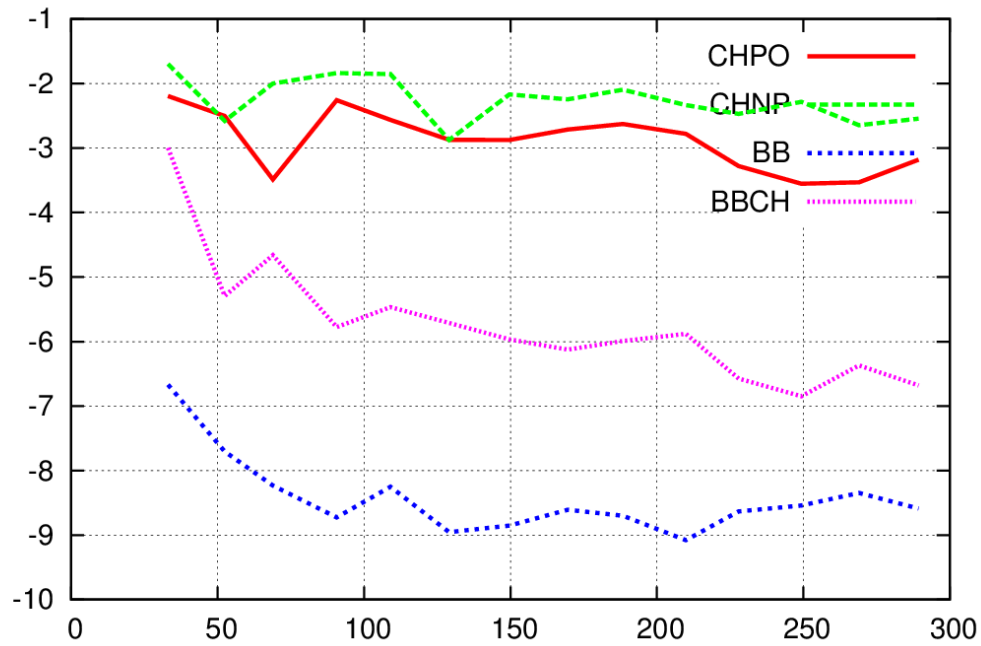
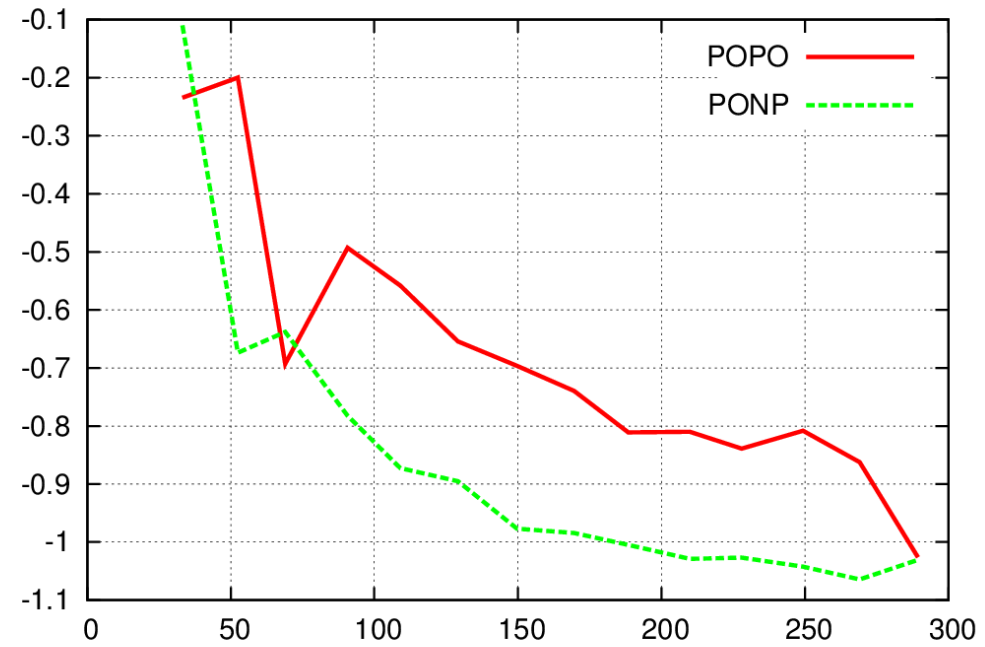
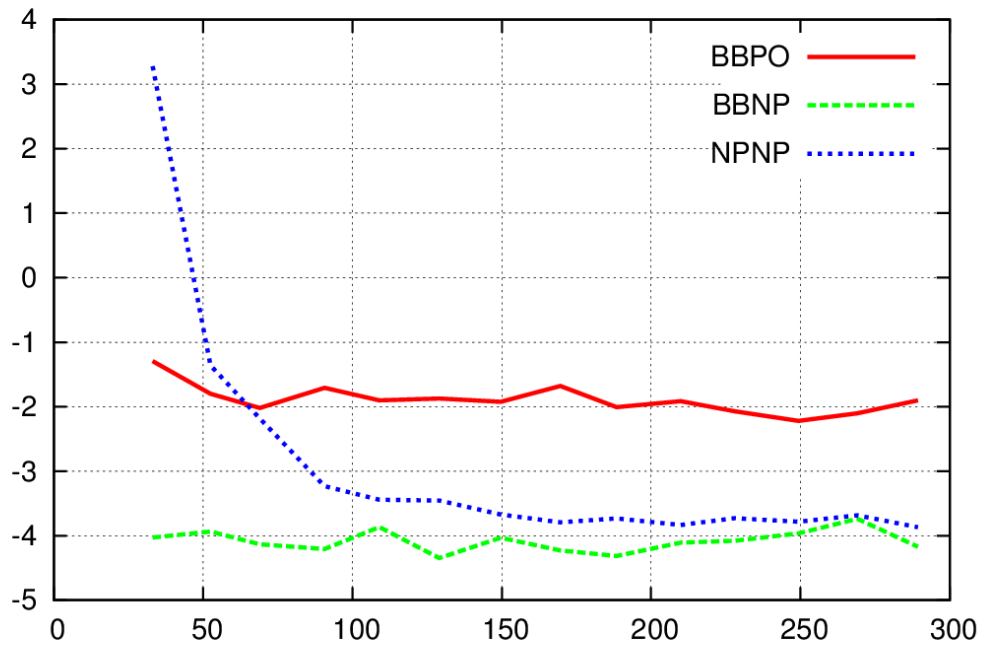
ASP



Effect of protein **size** on average RIE values

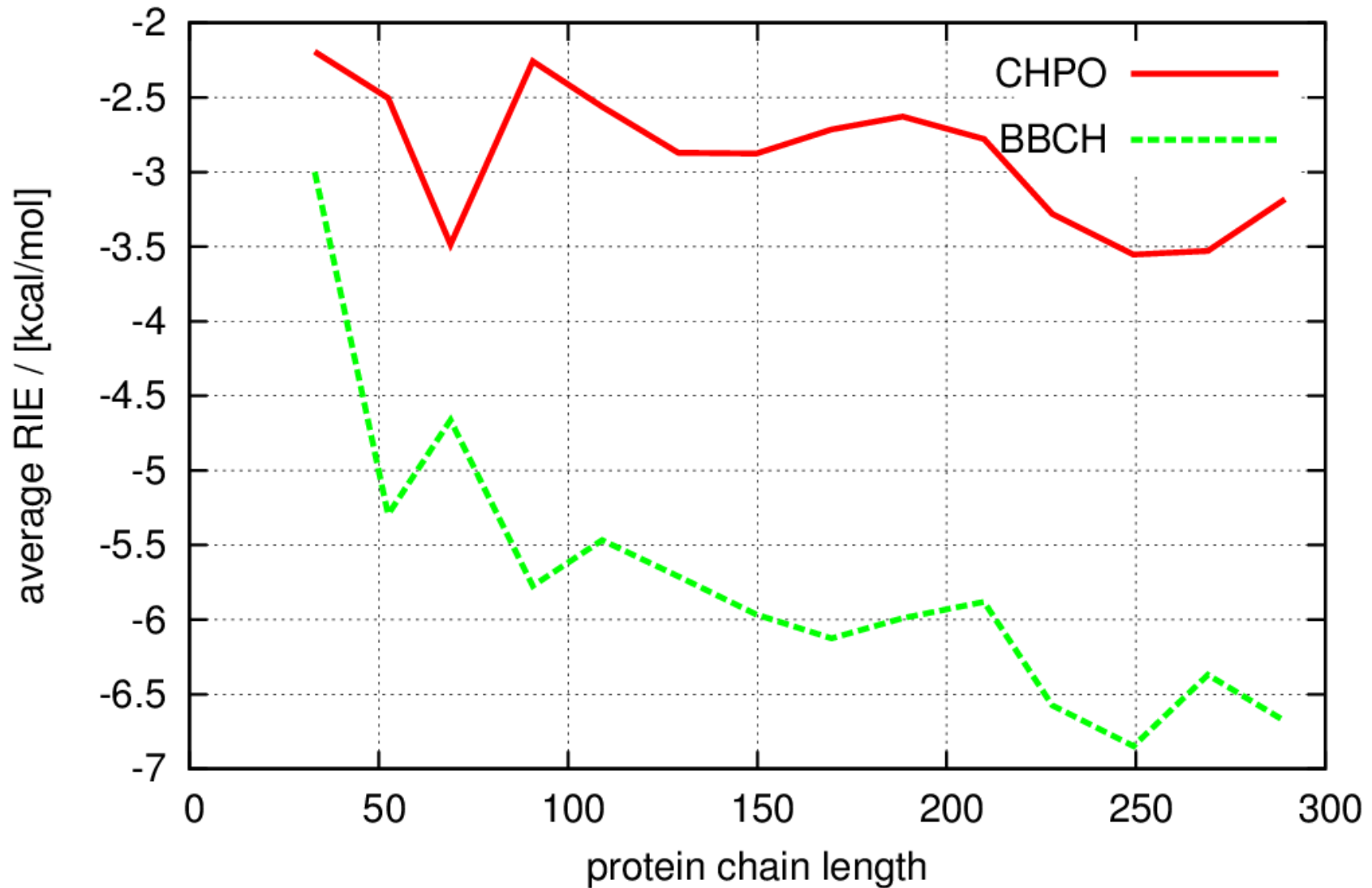


Size dependence of RIEs of certain type.
Averaged through whole structure set.



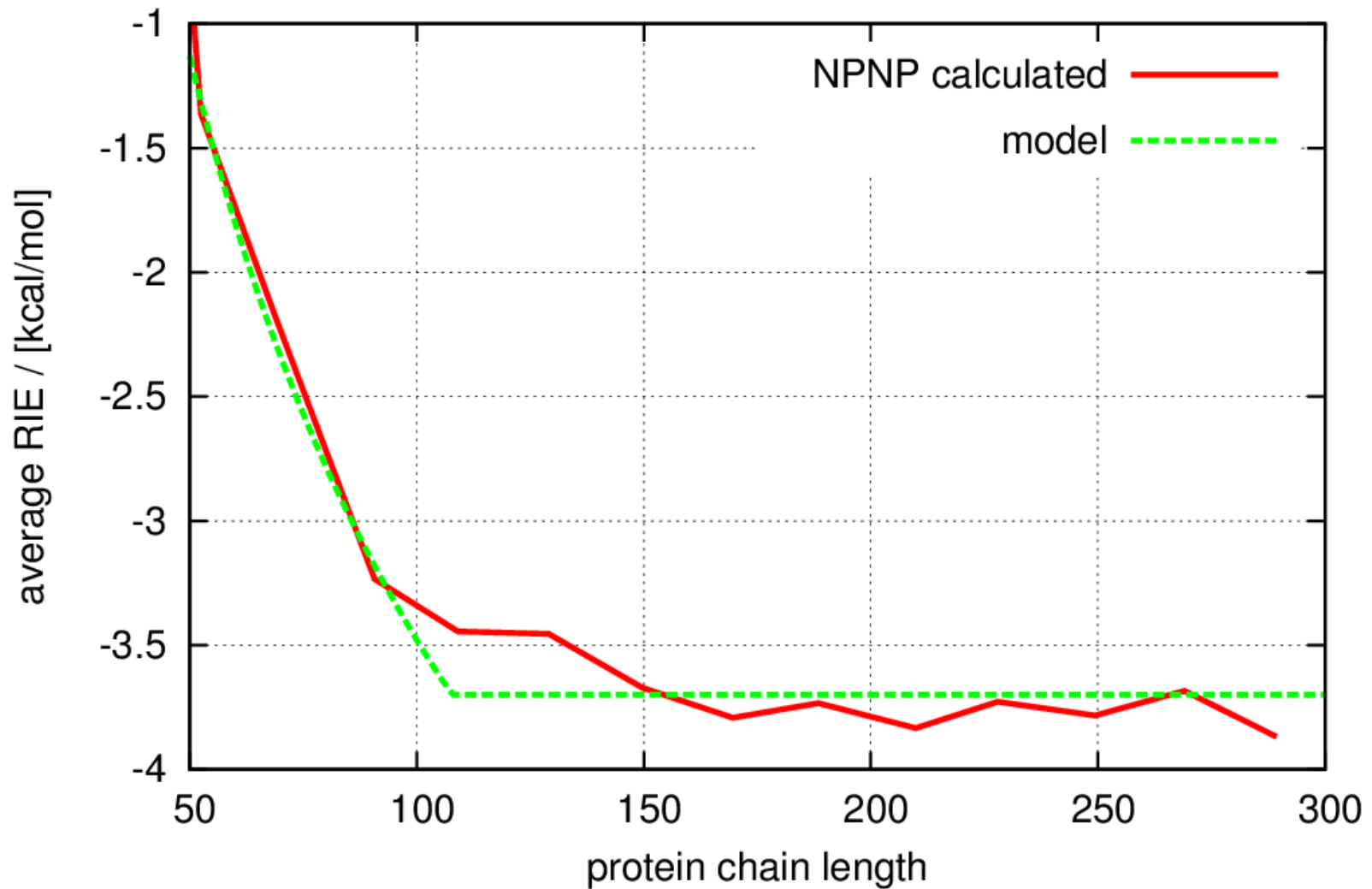
- CHPO and BBCH interactions are coupled

RIE distribution dependences of CHPO and BBCH on size.



$$E = \begin{cases} E = E_{\infty} \left(1 - kN^{-\frac{1}{3}}\right) & : N \leq N_D \\ E_D & : N > N_D \end{cases}$$

NPNP RIE dependence on size – comparison of model and calculated values.



Conclusions

- **Secondary structure content does not correlate** with RIE distributions except for BB RIEs.
- BB RIEs distributions have more than one peak and are different for each AA.
- Dependence of NPNP average RIEs can be used to determine **domain size** about 108 residues.
- BB and PO interactions are coupled.

Dr. Jiří Vondrášek

Jiří Vymětal

Jiří Kysilka

Karel Berka



Prof. Pavel Hobza
and the Group

This work was supported by Grant No. P208/10/0725 from the Czech Science Foundation, Grants LH11020 and LC512 from the Ministry of Education, Youth and Sports of the Czech Republic.