Intramolecular Interactions in Globular Proteins

Boris Fačkovec

Advisor: RNDr. Jiří Vondrášek, CSc.
Katedra fyzikální a makromolekulové chemie
Přírodovědecká fakulta UK

21st December 2011

Introduction - physical studies of proteins

Sequence-structure-stability-function relationships in globular proteins - central to biochemistry

- 1931 first theory of protein denaturation (Wu)
- 1958 statistical thermodynamics of polymers (Zimm)
- 1958 first resolved protein X-ray structure (Kendrew)
- 1961 Anfinsen's experiments
- 1977 first protein simulation (Karplus)
- 1983 Go model
- 80's-90's intensive development of force fields (Kollman, Jorgensen)
- 1985 knowledge-based force fields (Jernigan, Miyazawa)
- 90's - extensive calorimetric (DSC) studies (Privalov, Makhatadze)
- Denatured state investigations (Shortle)

Introduction - recent development

- 1993 sidechain atlas (Thornton)
- 90's lattice simulations (Shakhnovich)
- 1995 energy landscape perspective (Bryngelson, Onuchic)
- 1995 First force field decomposition of interactions (Lazaridis)
- 2000 Variational theory, spin glasses, principle of minimal frustration (Wolynes)

- 2005 stabilization of rubredoxin by strong dispersion interactions in its core (Vondrasek)
- 2008 identifying stabilizing residues by IEM calculations (Biedermannova)
- 2010 IEM development - fragmentation and QM calculations of protein molecules (Berka)
- 2010 FF calculations surprisingly good agreement with benchmark QM (Kolar)

Introduction - Interaction energy matrix (IEM)

- Fragmentation of a protein native structure
- QM calculations
- Classification of fragments - backbone BB, sidechain charged CH, polar PO, non-polar NP
- Pair additivity!

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

Figure: Interaction energy matrix of backbone-backbone interactions for a short peptide

'interaction' = inter-residual non-covalent interaction in single structure

Introduction - Types of intramolecular interactions

- Charged ion-ion
  - high values of IEs in IEMs not in correspondence with real stabilization effect
  - high compensation of attractive and repulsive interactions
  - high compensation of interactions in native and unfolded state
  - exceptionally high enthalpic-entropic compensation
- Charged multipoles
- Backbone
- van der Waals, stacking
  - short ranged - small compensation in
  - always attractive - small compensation
  - probably undervalued in IEMs
  - hydrophobic residues burial - folding driving force

Introduction - Objectives

- Characterization of magnitudes and distributions of inter-residual non-covalent interaction energies
- Development of unified *in silico* treatment - solution to problems with charged residues
- Decomposition of stabilizing energy

Methods

- Structure set selection - X-ray resolution <2 Å, single-stranded, no ligands $\rightarrow$ 1358 structures
- Optimization of hydrogen atoms - GROMACS, OPLS FF
- Fragmentation and IEM calculation $\rightarrow$ 10 IEMs
- Terminal backbones not considered,
- HIS double protonated $\rightarrow$ charged residue
- Only interactions with IE<-0.05 kcal/mol or IE>0.05 kcal/mol were sampled
- Subsets - size and secondary structure content

Results - Residue interaction energy distributions

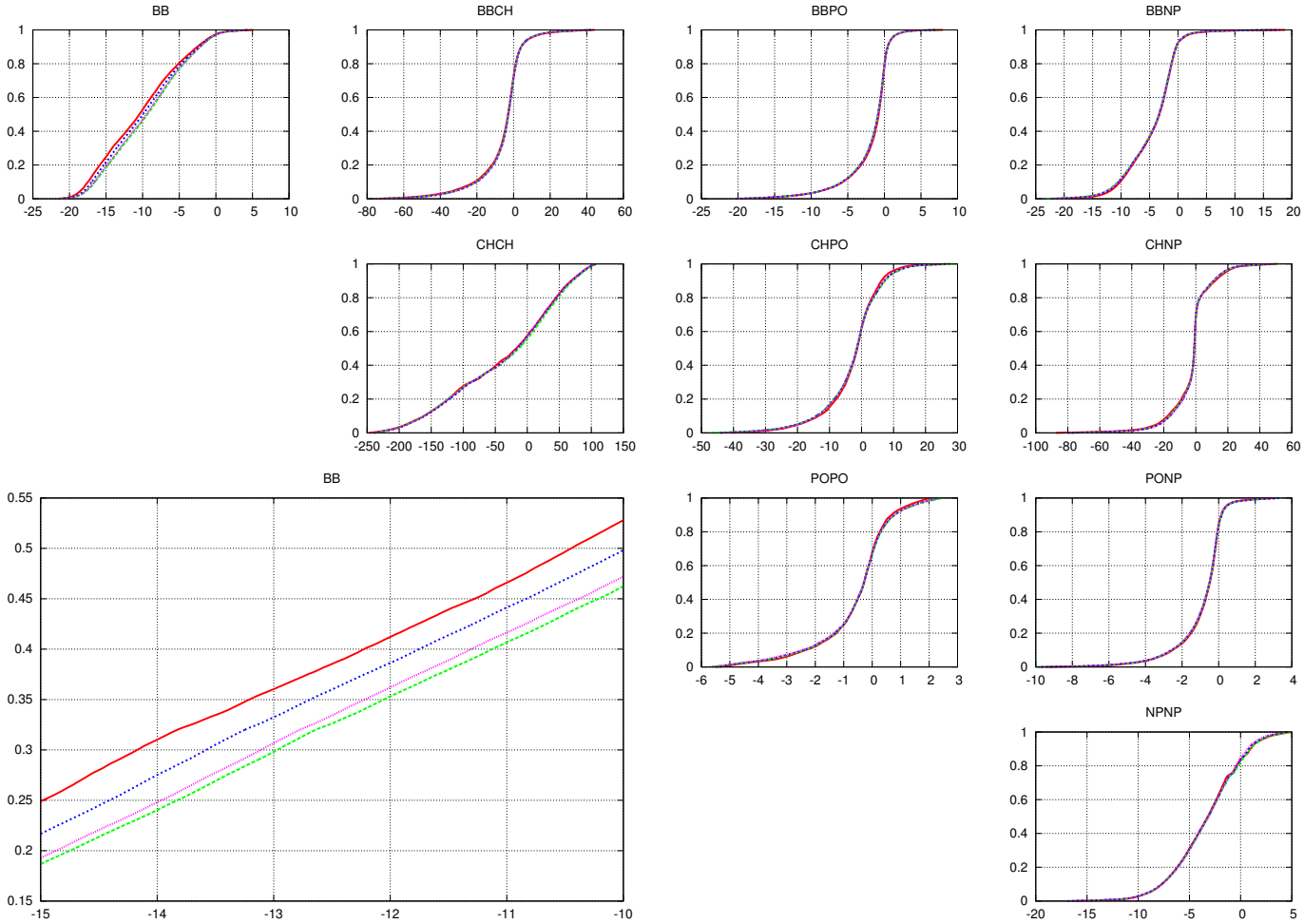

Figure: Distribution of RIE for all types of interactions. Various curves represent secondary structure particular classes.

Results - Domain size in globular proteins

$$E = \begin{cases} E_D(1 - kN^{\frac{1}{3}}) & N \leq N_D \\ E = E_D & N \geq N_D \end{cases} \quad (1)$$

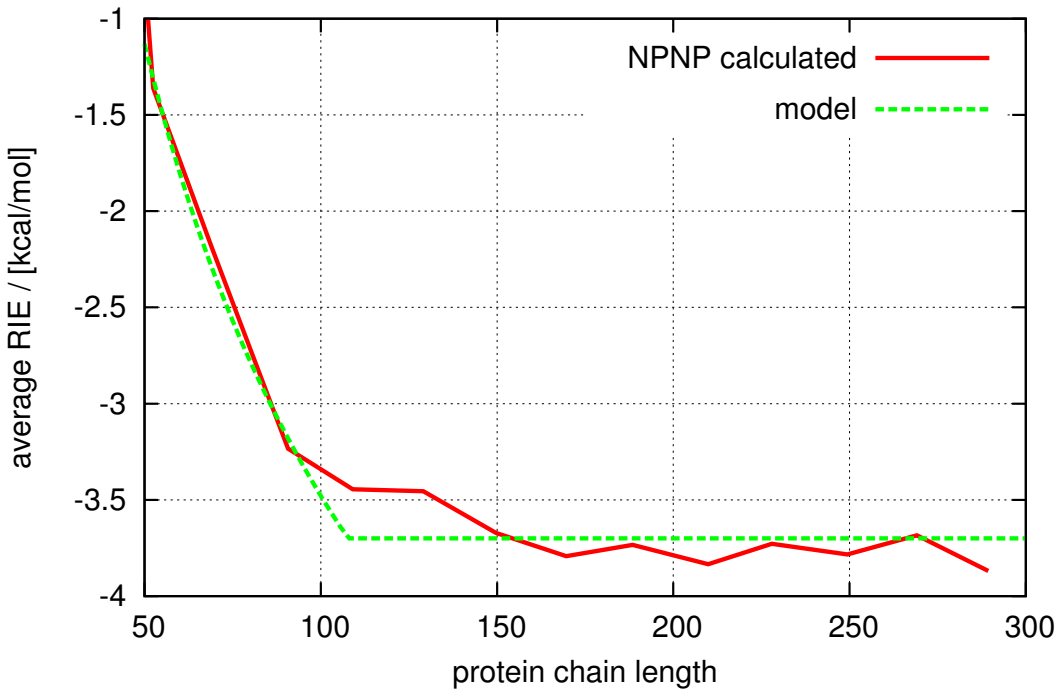

Figure: Average RIE - size dependence of NPNP interactions}HCIE of BB-BB interactions.

Results - Interaction energy distributions

- number of contacts increases immensely with decreasing IE definition, diverges to ∞ at IE=0
- Cumulative distribution of contributions to sum of IEs (DCIE) and its derivative (HCIE)

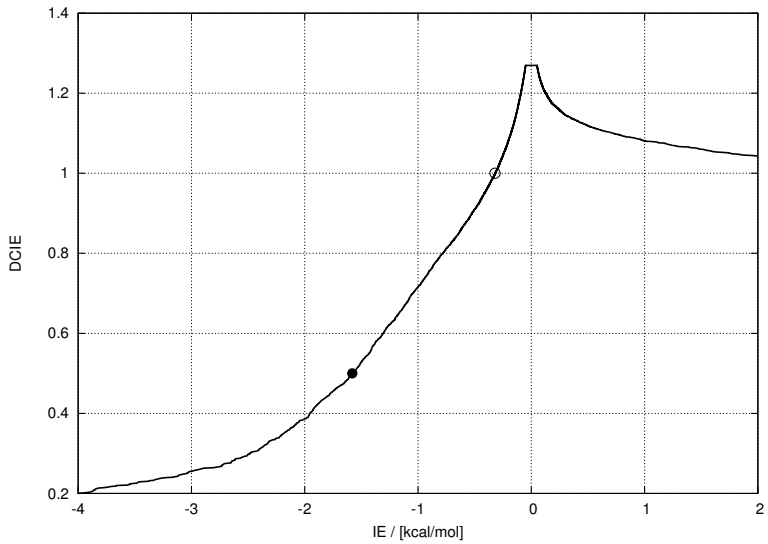

Figure: Cumulative distribution curve for IE of SER-TYR pair. BIE is the value of the interaction energy where the curve intersects 1 for the first time (-0.32 kcal/mol), $BIE_{0.5}$ is the value where it intersects 0.5 (-1.58 kcal/mol).

Results - Interaction energy distributions

Random energy model:

$$\mathrm{HCIE} = \mathrm{IE}\left(a_0 e^{-\left(\frac{\mathrm{IE}}{\sigma_0}\right)^2} + \sum_{i=1}^{t} a_i e^{-\left(\frac{\mathrm{IE}-\mathrm{IE}_i}{\sigma_i}\right)^2}\right) \quad (2)$$

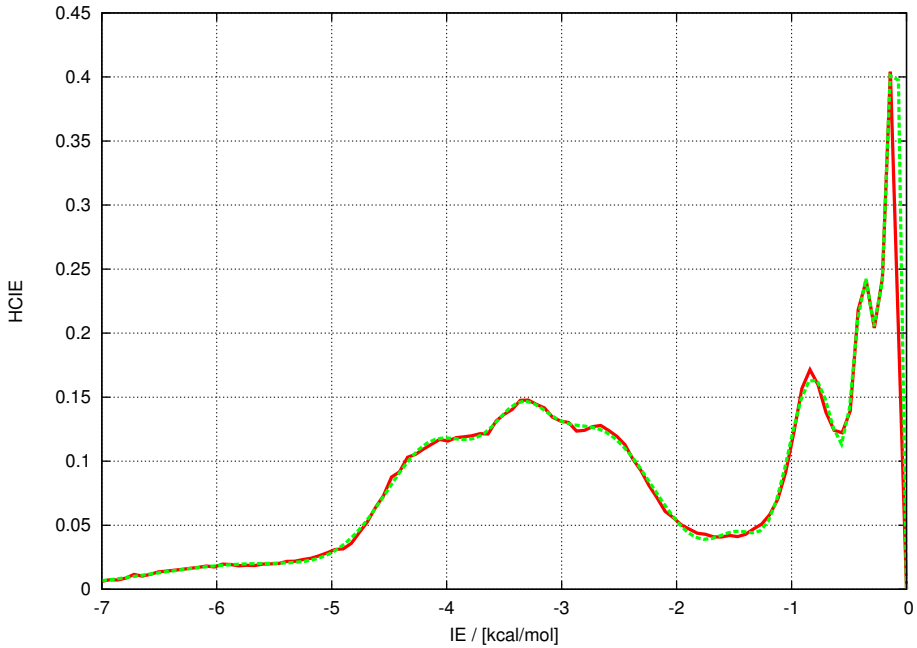

Figure: HCIE of BB-BB interactions. Red line represents calculated data, green line represents fit using 8 Gaussians.

Results - Optimum definition of interresidual contact based on interaction energy matrix calculations

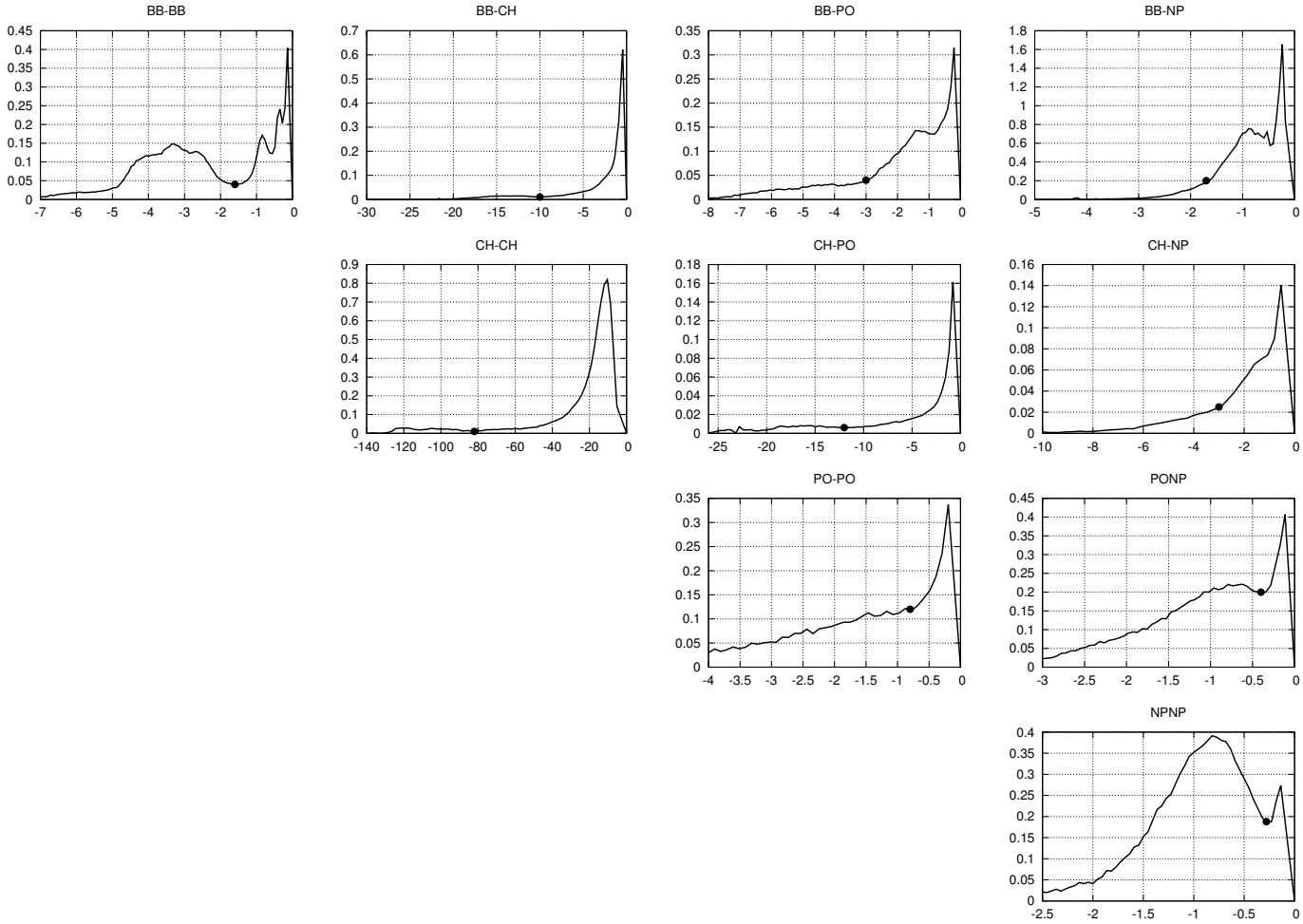

Figure: Contact definitions from HCIE curves for each type of interaction

Results - Optimum contact definitions and their protperties

Table: compens = ratio of the sum of all negative IEs to the sum of all IEs. Columns 3–6 show the order contributions of a particular type of interaction to a particular fragment type. x(BIEx) is the ratio of the energy content of the productive and all the interactions.

IE type	CD	BBCO	CHCO	POCO	NPCO	BIE	compens	x(BIEx)
BBBB	-1.6	1.08				-0.28	1.14	0.72
BBCH	-10	0.09	0.32			-3.5	4.04	0.39
BBPO	-3	0.1		0.37		-0.4	1.26	0.28
BBNP	-1.8	0.07			0.15	-0.1	1.05	0.08
CHCH	-82		0.16			-69	11.7	0.81
CHPO	-12		0.1	0.11		-4	3.03	0.47
CHNP	-3		0.16		0.1	-1.37	2.5	0.44
POPO	-0.8			0.33		-0.5	1.29	0.9
PONP	-0.4			1.44	0.82	-0.1	1.06	0.82
NPNP	-0.3				1.48	-0.19	1.09	0.96

Conclusion

- RIE characterized by magnitudes and distributions
- No correlation between sidechain IEs and secondary structure content
- Typical one-domain protein length - 110 residues
- Random energy model can be very successfully applied for IE statistics
- Compensation of positive and negative interactions characterized
- Contact definitions for each type of IEs $\rightarrow$ contact orders 1.34 for BB, 0.74, 2.25 and 2.56 for the CH, PO and NP sidechains

Future directions

- Web application on IOCB's site - IEM calculation, contact matrix
- Hydrophobic core definition - clusters in contact graphs
- Scaling factors $\rightarrow$ energy functions

References

- "Decomposition of Intramolecular Interactions Between Amino-Acids in Globular Proteins - A Consequence for Structural Classes of Proteins and Methods of Their Classification", **Fackovec B** and Vondrasek J, 2011, chapter 20, "Systems and Computational Biology - Molecular and Cellular Experimental Systems"

- "Optimal definition of inter-residual contact in globular proteins based on pairwise interaction energy calculations", **Fackovec B** and Vondrasek J, Bioinformatics, submitted

Acknowledgement

Jiří Vondrášek Jiří Vymětal Karel Berka
Jiří Kysilka