Atlas of Side-chain Interactions
The Beginning

Evaluation of the interaction energy between amino acid pairs in the most populated side-chain contacts in proteins by means of *ab initio* calculations.

K. Berka[1,2], J. Vondrášek[1], R. Laskowski[3], P. Hobza[1]

1) Institute of Organic Chemistry and Biochemistry, Centre for Complex Molecular Systems and Biomolecules, Flemingovo nam. 2, Prague 6, 166 10, Czech Republic
2) Charles University, Department of Physical and Macromolecular Chemistry, Albertov 6, Praha 2, 128 43, Czech Republic.
3) EMBL Outstation - Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
E-mail: karel.berka@marge.uochb.cas.cz

ÚOCHB AV ČR

EMBL-EBI
European Bioinformatics Institute

Introduction

The Protein Side-Chain Atlas [1] shows the most common 3D contacts between the amino acids within the known protein structures. The atlas is based on the printed atlas of **Singh & Thornton** [2]. It presents all **400** possible pairwise interactions between the **20** amino acid side-chains.

We utilize up to date version of protein side-chain atlas to determine the most populated amino acid interactions in proteins. We evaluated interaction energy for chosen 32 pairs including the most populated as well as representative interactions by means of a high level *ab initio* calculations to get the first rough estimate of potential energy surface.

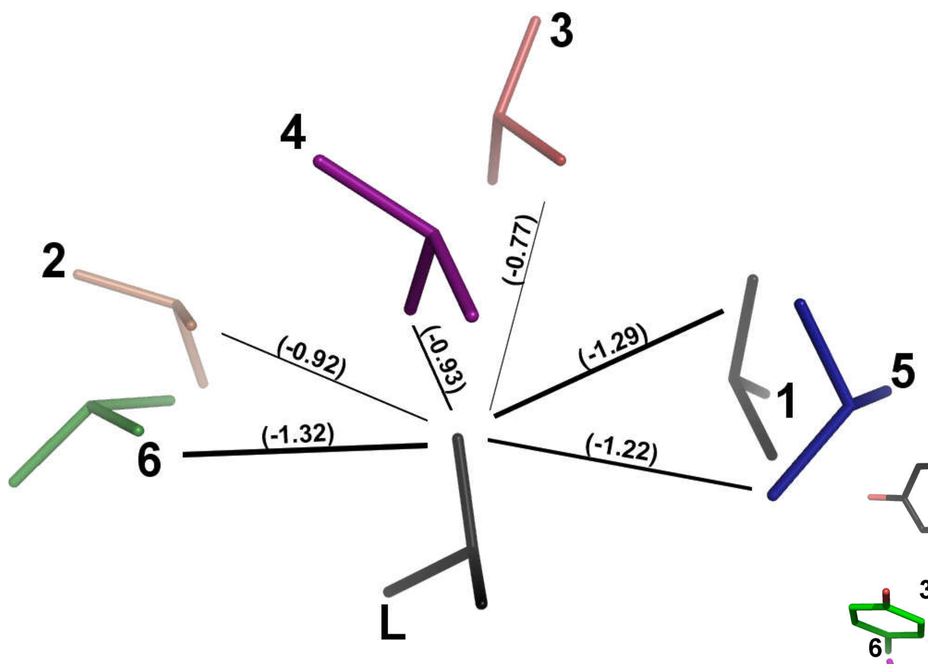

Figure 1 – The most populated side-chain contacts Leu-Leu with interaction energies.

Results

The most populated side-chain contacts are in Table 1.

Their interaction energies are all stabilizing in Table 2.

The interaction energy range (See Table 3 and Figure 1,2) is considerable for interacting polar residues, while in the case of non-polar amino acids is much smaller.

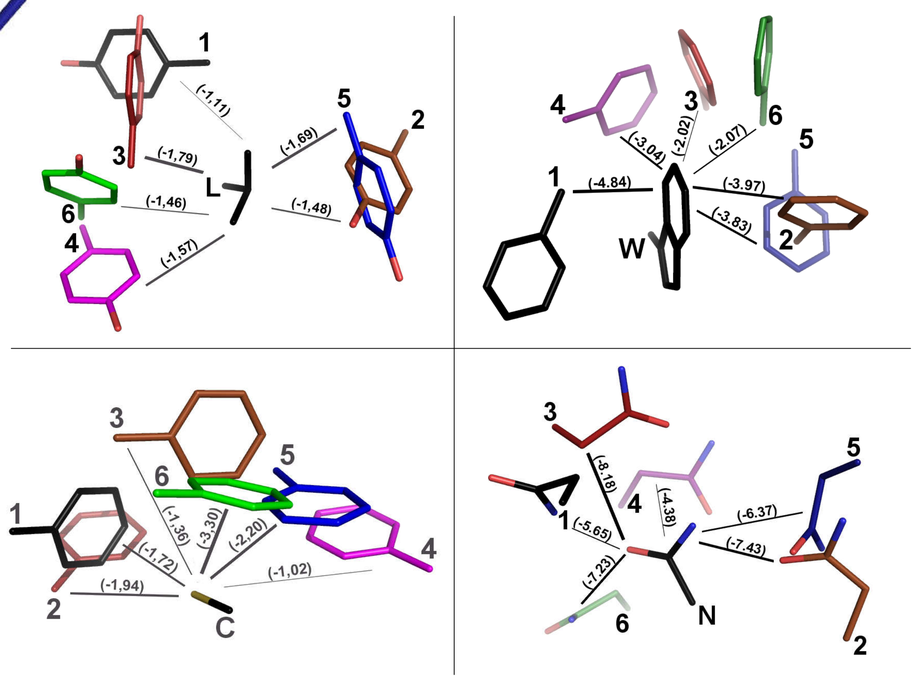

Figure 2 – Different types of side-chain contacts: aliphatic-aromatic (Leu-Tyr); aromatic-aromatic (Trp-Phe); polar-aromatic (Cys-Phe); polar-polar (Gln-Asn).

	Central aminoacid	Interacting aminoacid	Contacts in Atlas
1	Leu	Leu	**47638**
2	Val	Leu	**27218**
3	Leu	Val	**27217**
4	Ile	Leu	**26652**
5	Leu	Ile	**26651**
6	Val	Val	**19723**
7	Ile	Ile	**18624**
8	Ala	Leu	**15282**
9	Leu	Ala	**15280**
10	Tyr	Leu	**12039**
11	Leu	Tyr	**12030**
12	Phe	Phe	**11127**
13	Thr	Leu	**8236**
14	Leu	Thr	**8233**
15	Leu	Met	**6993**
16	Met	Leu	**6993**
17	Phe	Tyr	**6660**
18	Tyr	Phe	**6658**
19	Leu	Trp	**6487**
20	Trp	Leu	**6485**
21	Tyr	Tyr	**5179**
22	Phe	Trp	**3676**
23	Trp	Phe	**3676**
24	Ser	Thr	**3132**
25	Thr	Ser	**3132**
26	Asn	Gln	**2227**
27	Gln	Asn	**2217**
28	Cys	Phe	**1987**
29	Phe	Cys	**1987**
30	Met	Met	**1973**
31	Cys	Met	**641**
32	Met	Cys	**641**

Table 1 – The most populated side-chain contacts of several types of interaction.

Cβ	GLY	ALA	VAL	ILE	LEU	PHE	TYR	TRP	HIS	PRO	THR	SER	ASN	GLN	CYS	MET	LYS	ARG	ASP	GLU
GLY																				
ALA					-0.44															
VAL			-0.76		-0.87															
ILE				-1.00	-1.03															
LEU		-0.63	-0.90	-1.09	-1.08		-1.51	-2.92			-0.76					-1.22				
PHE						-2.41	-2.80	-2.73							-1.59					
TYR					-2.33	-3.12	-2.68													
TRP					-2.77	-3.29														
HIS																				
PRO												-2.79								
THR					-0.75						-2.90									
SER													-6.54							
ASN														-6.04						
GLN						-1.92										-1.92				
MET					-1.46										-2.45	-2.13				
LYS																				
ARG																				
ASP																				
GLU																				

Legend
-0 – -0.99
-1 – -1.99
-2 – -2.99
-3 – -3.99
-4 – -4.99
Above 5

Table 2 – Average interaction energies for side-chain pairs. All energies are in kcal/mol.

Conclusions

- Interaction energy for 32 pairs of interacting side-chains was determined at the highest *ab initio* level in the gas phase.
- All calculated energy contributions were stabilizing, pointed out very important and we guess *general* feature of amino acid side chains internally packed in proteins.
- Our ultimate goal is to determine computationally complete 20x20 matrix of amino acid pair interactions in a gas phase and in a solvent environment for further bioinformatics utilization.

Materials and Methods

Interaction energy evaluation

The heavy atoms coordinates for side-chain beginning at Cβ were kept fix at the positions from the X-ray structures for the cluster representative from the side-chain atlas. Position of the hydrogens was then optimized at the DFT/TPSS/TZVP level. The pair stabilization energies were determined at the CCSD(T) level with complete basis set (CBS) of atomic orbitals. The CCSD(T)/ CBS interaction energy was approximated as:

$$\Delta E^{CCSD(T)}|_{CBS} = \Delta E^{MP2}|_{CBS} + (\Delta E^{CCSD(T)} - \Delta E^{EMP2})|_{small\ basis\ set} \quad (1)$$

The former term was determined using Helgaker [3,4] extrapolation scheme.

The Hartree-Fock and correlation MP2 energies necessary for the extrapolation were determined with aug-cc-pVXZ (X=D, T) basis sets. The CCSD(T) term was calculated with smaller basis set, 6-31G*(0.25). The use of smaller basis set is based on the fact that the difference between the MP2 and CCSD(T) interaction energies (contrary to MP2 and CCSD(T) interaction energies themselves) is much less dependent on the size of the basis set and the 6-31G*(0.25) basis set already gives satisfactory values of this difference.

All interaction energies were corrected for the basis set superposition error.

Central aminoacid	Interacting aminoacid	Interaction energies (kcal/mol) Average	Minimal	Maximal	Standard deviation
Asn	Gln	**-6.54**	-4.38	-8.18	1.25
Gln	Asn	**-6.04**	-3.93	-7.19	1.15
Trp	Phe	**-3.29**	-2.02	-4.84	1.03
Tyr	Phe	**-3.12**	-2.58	-3.68	0.38
Leu	Trp	**-2.92**	-0.82	-4.01	1.00
Ser	Thr	**-2.90**	-0.27	-5.49	2.13
Thr	Ser	**-2.79**	-0.63	-4.90	1.76
Phe	Tyr	**-2.77**	-2.12	-3.52	0.48
Trp	Leu	**-2.77**	-1.16	-4.14	1.06
Phe	Trp	**-2.73**	-1.52	-4.58	1.00
Tyr	Tyr	**-2.68**	-1.89	-3.78	0.75
Met	Cys	**-2.45**	-1.27	-3.12	0.61
Phe	Phe	**-2.41**	-1.85	-2.91	0.39
Tyr	Leu	**-2.33**	-1.07	-3.15	0.66
Met	Met	**-2.13**	-1.40	-2.54	0.39
Cys	Phe	**-1.92**	-1.02	-3.30	0.72
Cys	Met	**-1.92**	-1.30	-3.07	0.61
Phe	Cys	**-1.59**	-0.32	-3.89	1.21
Leu	Tyr	**-1.51**	-1.11	-1.79	0.22
Met	Leu	**-1.46**	-0.75	-1.96	0.41
Leu	Met	**-1.22**	-0.97	-1.88	0.31
Leu	Ile	**-1.09**	-0.71	-1.39	0.22
Leu	Leu	**-1.08**	-0.77	-1.32	0.21
Ile	Leu	**-1.03**	-0.46	-1.79	0.40
Ile	Ile	**-1.00**	-0.70	-1.21	0.21
Leu	Val	**-0.90**	-0.63	-1.13	0.15
Val	Leu	**-0.87**	-0.51	-1.12	0.20
Leu	Thr	**-0.76**	-0.69	-0.92	0.08
Val	Val	**-0.76**	-0.55	-1.00	0.18
Thr	Leu	**-0.75**	-0.51	-0.99	0.16
Leu	Ala	**-0.63**	-0.51	-0.81	0.10
Ala	Leu	**-0.44**	-0.12	-0.71	0.18

Table 3 – Variation of interaction energies for side-chain pairs. All energies are in kcal/mol.

References

[1] (http://www.biochem.ucl.ac.uk/bsm/sidechains/)

[2] Singh J, Thorton JM, Atlas of Protein Side Chain Interactions, IRL PRESS, OXFORD, UK, 1992

[3] Halkier A, Helgaker T, Klopper W, *Chem. Phys. Lett*. 286, 243 (1998)

[4] Halkier A, Helgaker T, Jorgensen P, *Chem. Phys. Lett*. 302, 437 (1999)

Acknowledgement

This work was supported by grants 203/05/0009, 203/06/1727 and 203/05/H001 from the Grant Agency of the Czech Republic, grant A400550510 from the Grant Agency of the Academy of Sciences of the Czech Republic, and grant LC512 Ministry of Education (MSMT) of the Czech Republic. It was also part of research projects Z40550506 and MSM6198959216.