Ars Separatoria Acta 6 (2008) 7-14

ARS SEPARATORIA ACTA
www.arsseparatoriaacta.com

SELF-CONCERTED HYDROPHOBICITY SCALE BASED ON THE ACDLOGP OF A COMBINATION OF 8000 TRIPEPTIDES AND ITS APPLICATION FOR IDENTIFICATION OF PROTEIN ACTIVE SITES

Krystian EITNER[1)], Uwe KOCH[2)], Tomasz GAWĘDA[3)], Stephane SPIESER[4)], Amadeusz FORNALIK[5)], Marcin HOFFMANN[1,3)]

[1)]Adam Mickiewicz University, Faculty of Chemistry, Grunwaldzka 6, PL-60-780 Poznań, Poland
[2)]Lead Dicovery Center, Emil-Figge-Str. 76a, 44227 Dortmund, Germany
[3)]BioInfoBank Institute, Św. Marcin 80/82 lok. 355, 61-809 Poznań, Poland
[4)]Department of Chemistry, Istituto di Ricerche di Biologia Molecolare P. Angeletti, Merck Research Laboratories, Rome, Via Pontina km 30600, Pomezia 00040, Italy
[5)]Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa, Rua Ernesto de Vasconcelos, C8, Campo Grande, 1749-016 Lisboa – Portugal

ABSTRACT

Partition coefficients, expressed as logP, were calculated using Advanced Chemistry Development software (ACDlogP) [1] of all combinations of three amino acids (8000 tripeptides). Using our proprietary script, we generated a combination of 8000 tripeptides of 20 amino acids in the FASTA format and, subsequently, full atom Cartesian coordinates were generated. The ACDlogP's of the generated tripeptides were calculated. Using the Kyte-Doolittle amino acid hydrophobicity scale [2, 3], the value of the correlation coefficients with the calculated ACDlogP values was determined. Hydrophobicity values were assumed as the arithmetic mean of the hydrophobicity of the three amino acids in the tripeptide. Optimisation of the theoretical hydrophobicity scale by minimisation of the correlation coefficient between the calculated ACDlogP values and the hydrophobicity for the tripeptides provided amino acid hydrophobicity; on this basis, the amino acids were divided into 7 groups. The new scale was normalised and implemented using the fuzzy-oil-drop model method to determine the theoretical protein active site of the 1HCK protein based on lipophilic hot spots on the protein surface. The results were compared with the respective results for the Kyte-Doolittle scale and the actual active site with ATP as the ligand.

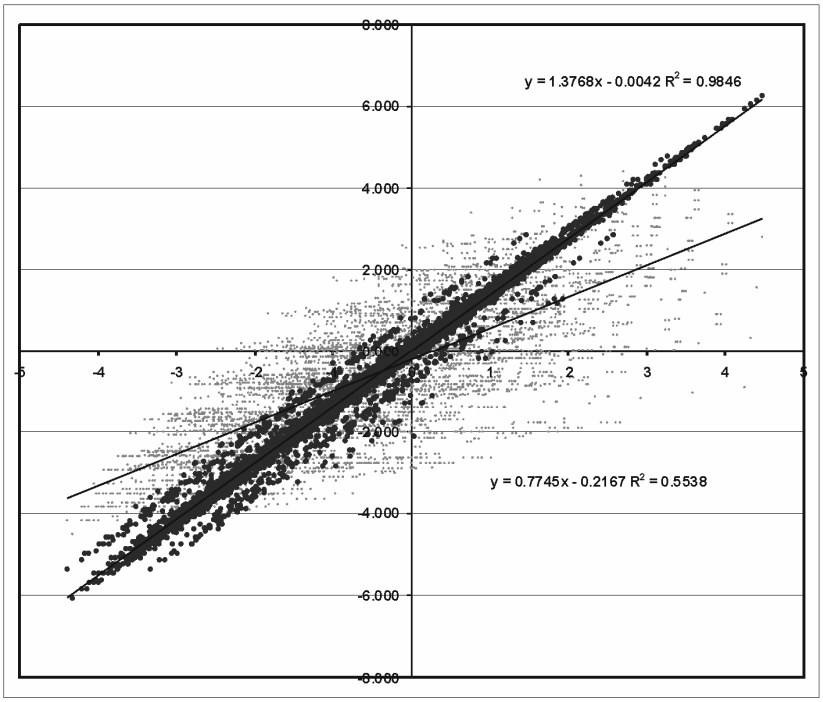
Keywords: active site identification, fuzzy-oil-drop model, hydrophobicity, tripeptides

INTRODUCTION

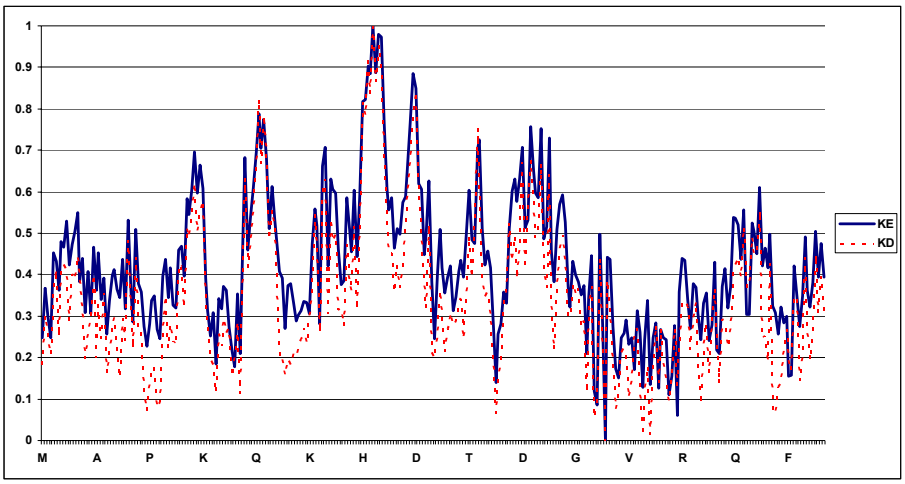
The quantitative measure of the hydrophobicity of amino acids is their partition coefficient between the aqueous and lipid phases. In model experimental research based on the widely used experimental model the lipid phase is replaced with 1-octanol. Since the 1950's and Kauzmann's work [2], hydrophobicity is considered the main driving force for the formation of tertiary and quaternary protein structures. According to the theory, hydrophobic amino acid residues cause the protein to fold so that its surface accessible to the aqueous environment is as hydrophilic as possible. Owing to this, the resulting structure forms the active site. The active site is usually a cavity which is significantly more lipophilic then the surrounding surface. This mechanism for the formation of protein tertiary structures is critical for transmembrane proteins while in the case of globular proteins this formation mechanism ensures both good protein solubility in bioelectrolytes and the existence of a formed and functional protein active site. Based on the knowledge about hydrophobicity distribution in protein structures, the amino acids involved in binding inhibitors and proper substrates can also be deduced.

Calculation procedure

The full-atom structures of all possible 8000 tripeptides were obtained using FASTA and SDF formats. Subsequently, the ACDlogP physicochemical descriptor was calculated for the resulting tripeptides. The tripeptide hydrophobicity coefficients were assumed to be the arithmetic mean of the hydrophobicity coefficients of the respective amino acids according to the Kyte-Doolittle scale [3]. Determination of the correlation of these hydrophobicity coefficients with the calculated ACDlogP values for the tripeptides showed that the correlation coefficient was 55%. In order to optimise the hydrophobicity scale, the mean square deviations of the Kyte--Doolittle scale correlation was minimised, varying the amino acid hydrophobicity coefficients. To this end, the Excel Microsoft Office solver was employed. The resulting scale has 98% correlation with the ACDlogP values (Plot 1).



Plot 1. Relationship between linear combination of Kyte-Doolittle scale and ACDlogP's of tripeptides (pink), and between optimised amino acid hydrophobicity scale and ACDlogP (blue).



Plot 2. Distribution of 1HCK protein hydrophobicity using Kyte-Doolittle scale (KD, red) and ACDlogP optimisation-based scale (KE, dark blue).

Based on the scale, the amino acids were divided into seven groups of varying hydrophobicity (Table 1).

Table 1. Kyte-Doolittle scale and optimised amino acid scale.

AA	K-D scale	Theor. scale	Theor. scale norm.
R	**– 4.50**	**– 6.07**	**1.000**
Q	**– 3.50**	**– 5.39**	**0.945**
N	**– 3.50**	**– 4.90**	**0.906**
S	– 0.80	– 4.25	0.852
E	– 3.50	– 3.93	0.827
H	– 3.20	– 3.72	0.810
K	**– 3.90**	**– 2.91**	**0.744**
G	**– 0.40**	**– 2.87**	**0.740**
T	**– 0.70**	**– 2.79**	**0.734**
P	– 1.60	– 1.51	0.630
A	1.80	– 1.41	0.622
D	– 3.50	– 0.90	0.581
M	**1.90**	**1.85**	**0.358**
V	**4.20**	**2.27**	**0.323**
C	**2.50**	**2.36**	**0.316**
Y	**– 1.30**	**3.17**	**0.251**
I	4.50	4.50	0.142
L	3.80	4.50	0.142
W	**– 0.90**	**5.94**	**0.026**
F	**2.80**	**6.26**	**0.000**

Scale evaluation using the Fuzzy-Oil-Drop model and comparison of results for the Kyte-Doolittle scale.

FUZZY-OIL-DROP MODEL AND CAVITY SPACE IDENTIFICATION

The Fuzzy-oil-drop model is based on the three-dimensional Gauss function distribution of hydrophobicity coefficient in the grid space of a protein structure [4, 5]. The area critical for the biological function of protein molecules is indicated by the irregularity of hydrophobicity distribution for the surface (calculated hydrophobicity coefficient higher than expected), and the cavity space (calculated hydrophobicity coefficient lower than expected). Such a model localizes the area responsible for ligand binding or protein-protein complex formation, based on the characteristics of spatial distribution of hydrophobicity in a protein molecule. Specific folding of proteins disturbing the regularity of the hydrophobicity coefficient distribution seems

to be a good indicator of structural differentiation for every individual protein. The regions recognised by high hydrophobicity density differences seem to reveal functionally important sites in proteins [5].

1HCK (HUMAN CYCLIN-DEPENDENT KINASE 2) [6] protein was selected for further tests, because its 3D structure with the natural ligand (ATP) in the active site has been determined crystallographically and thus all amino acids which surround the ligand at a radius of 4 Å can be determined. It is also important that 1HCK is a single chain protein, which further facilitates computational analysis.

The results show it clearly that the hydrophobicity scale enables identification of the protein active site. It is noted that due to the high mobility of the 10-18 terminal amino acids, results obtained based on the proposed model did not reveal the participation of the amino acids in ligand binding, even though their distance from the ATP ligand in the crystal structure is shorter than 4 Å (Figure 1). The theoretical scale we proposed and the calculations performed based on the Kyte-Doolittle coefficient indicated that 7 mutually remote peptide motives are involved in the formation of the active site cavity. The theoretical scale proves that optimisation of the 3D descriptor (ACDlogP) enables identification of protein active sites designed by homology and supports the interpretation of experimental data [7]. The Figures 2-5 presents graphical visualization of differences between the theoretical and Kyte-Doolittle scale results.

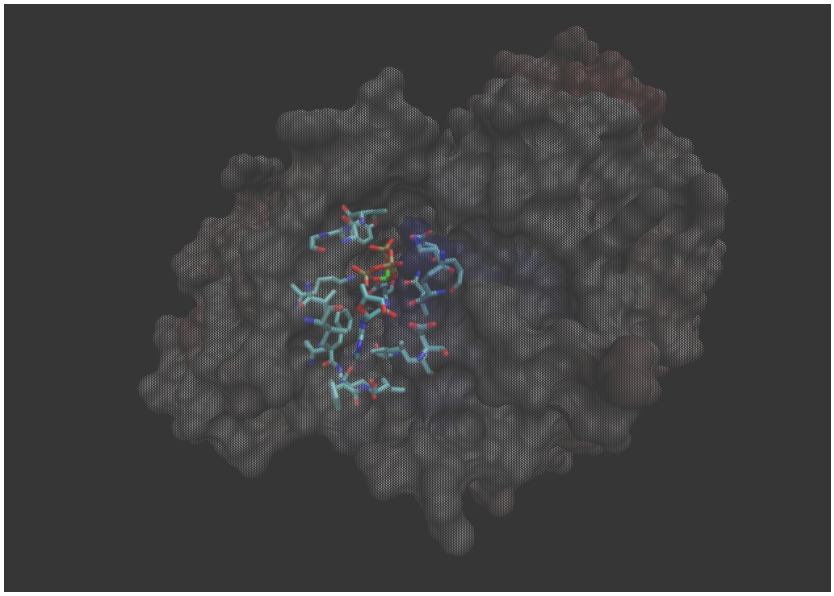


Fig. 1. 4 Å environment of protein active site around ATP.

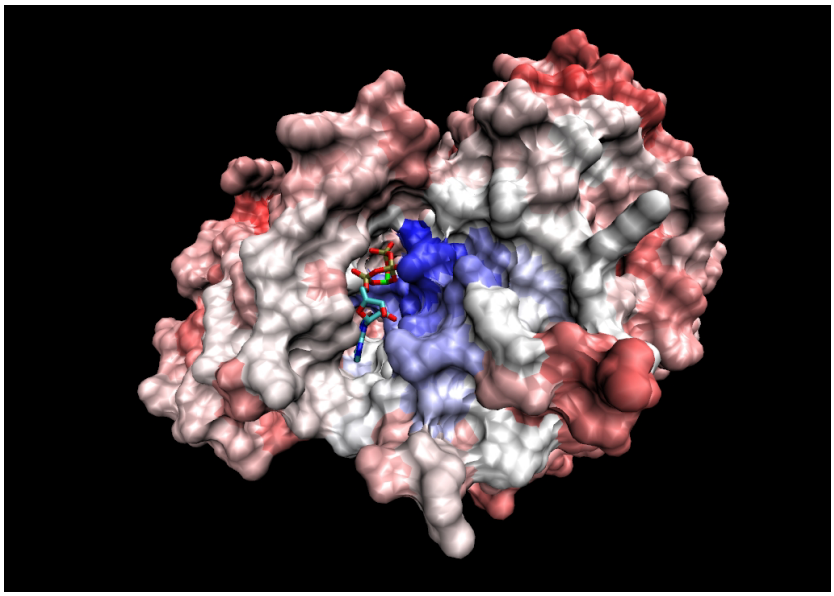


Fig. 2. Theoretically determined protein active site (blue) according to Kyte-Doolittle scale.

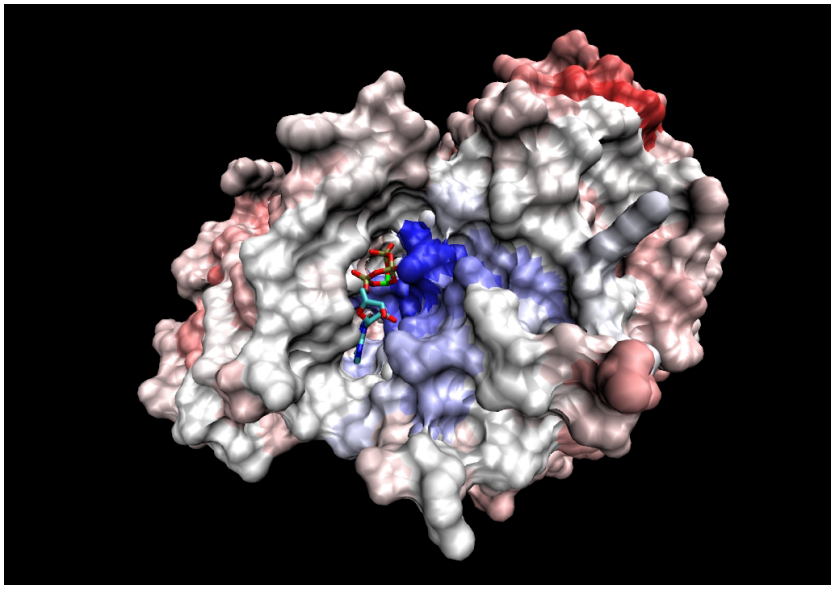


Fig. 3. Theoretically determined protein active site (blue) according to theoretically optimised scale.

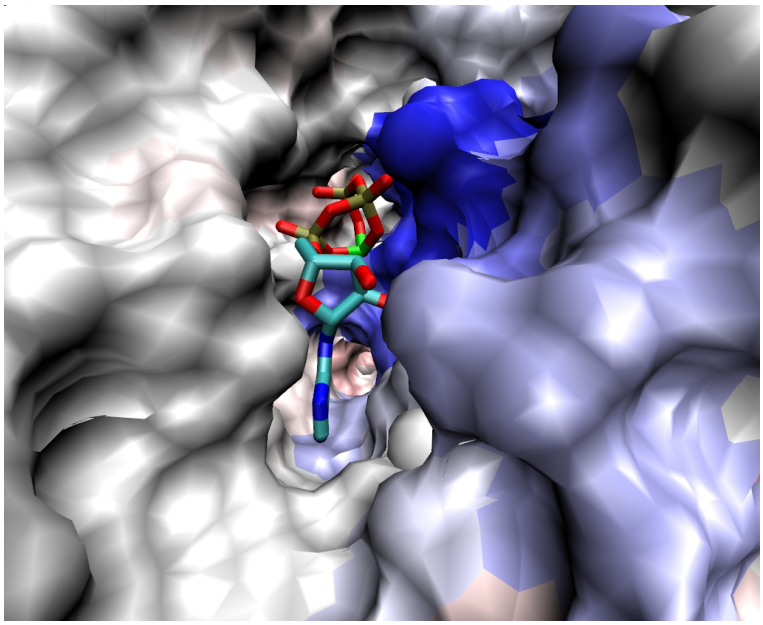


Fig. 4. Theoretically determined protein active site (blue) according to Kyte-Doolittle scale.

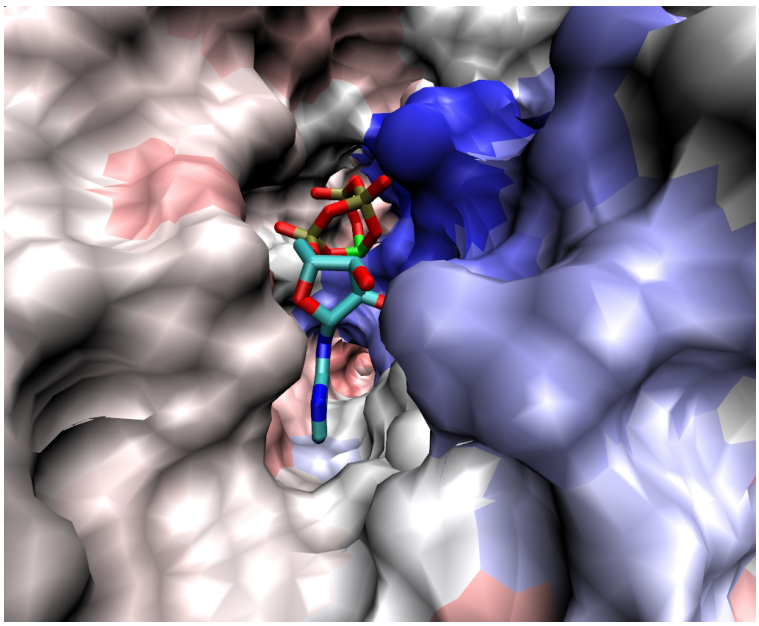


Fig. 5. Theoretically determined protein active site (blue) according to theoretically optimised scale.

SUMMARY

Protein active site analysis and identification is a complex and sophisticated task also due to the fact that not all proteins can be crystallographically analysed. On the other hand, the current homology-based protein design methods not always provide information about active sites. This paper shows how active and allosteric sites can be successfully identified using the Fuzzy-Oil-Drop model and a scale based on the ACDlogP's of 8000 tripeptides. The new scale was developed following optimisation of the correlation between ACDlogP values and hydrophobicity values of tripeptides on the hydrophobicity scale of three amino acid combinations using a topological descriptor. The theoretical scale is fully compatible with the experimental Kyte-Doolittle hydrophobicity scale, and analysis of the 1HCK protein proved correct the model compared with the analysis of the crystallographically characterised protein active site.

ACKNOWLEDGEMENT

KE and MH would like to thank the Foundation for Polish Science for the support through the FOCUS programme.

REFERENCES

[1] ACD/Log*P* DB. www.acdlabs.com/logp/ Advanced Chemistry Development, Inc., Toronto, ON, Canada. 15 January, 2007.
[2] Kauzmann W., 1959. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14, 1-63.
[3] Kyte J., Doolittle R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.
[4] Levitt M., 1976. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 104, 59-107.
[5] Brylinski M., Konieczny L., Roterman I., 2006. Hydrophobic collapse in (in silico) protein folding. Computational Biology and Chemistry 30, 255-267.
[6] de Bondt H.L., Rosenblatt J., Jancarik J., Jones H.D., Morgan D.O., Kim S.H., 1993. Crystal Structure of Cyclin-Dependent Kinase-2. Nature 363, 595.
[7] Leo A., Hansch C., Elkins D., 1971. Partition Coefficients and their uses. Chemical Reviews 6(71).