CROATICA CHEMICA ACTA
CCACAA **81** (2) 283–287 (2008)

ISSN-0011-1643
*CCA*-3242
*Original Scientific Paper*

# Modelling of ICAM-1 and LFA-1 Interaction Using Molecular Recognition Theory

**Nikola Štambuk,* Paško Konjevoda, Dražen Vikić-Topić, and Biserka Pokrić**

*Ruđer Bošković Institute, Bijenička cesta 54, HR-10002 Zagreb, Croatia*

RECEIVED AUGUST 13, 2007; REVISED FEBRUARY 24, 2008; ACCEPTED FEBRUARY 27, 2008

The model of ICAM-1 and LFA-1 molecular interaction was used to test the application of the Molecular Recognition Theory for the identification of several discontinuous binding regions, *i.e.* ligand-receptor sites, within large antigenic molecules. Molecular Recognition Theory is an applicable heuristic algorithm for identification of possibly interacting amino acid pairs in short linear epitope/paratope sites within larger molecules. However, in order to achieve better efficiency this heuristic algorithm of molecular recognition has to be combined to several other procedures: molecular hydropathy analyses, secondary structure prediction methods and protein database search. The limitation of the combined MRT-hydropathy analyses is in the fact that it cannot explain 3D protein interactions, but it can be a valuable starting point for a more complex computational and experimental analysis.

*Keywords*
molecular recognition
heuristic algorithm
ICAM-1
LFA-1
binding

## INTRODUCTION

Epitopes are regions on the antigen molecule to which antibody or T-cell receptor binds specifically.[1,2] Procedures that allow the identification of B-cell and T-cell epitopes on the protein are important for vaccine design and selective peptide immunomodulation.[2] Identification of epitope binding sites on the molecules is important for the immune system manipulations based on antibodies or short peptides.[2,3]

Molecular Recognition Theory (MRT) is thought to be an useful procedure for the modelling of peptide-receptor and antigen-antibody interactions.[4–10] It is based on an observation of Blalock *et al.*[4–10] that sense and antisense peptides, due their molecular hydropathy profile, have mutually complementary shapes which results in their interaction. In this paper we apply the Molecular Recognition Theory for the prediction of short linear segments of discontinuous binding regions of two large interacting molecules.

The analysed pair was human intercellular adhesion molecule 1 (ICAM-1) and its natural ligand leukocyte function-associated antigen-1 (LFA-1).[3]

## RESULTS AND DISCUSSION

### *Epitope Detection*

The analysis of the hydrophobic profile of human intercellular adhesion molecule 1 (ICAM-1) was performed by the algorithm of Kyte & Doolittle.[3,11] The first 27 ami-

* Author to whom correspondence should be addressed. (E-mail: stambuk@irb.hr)

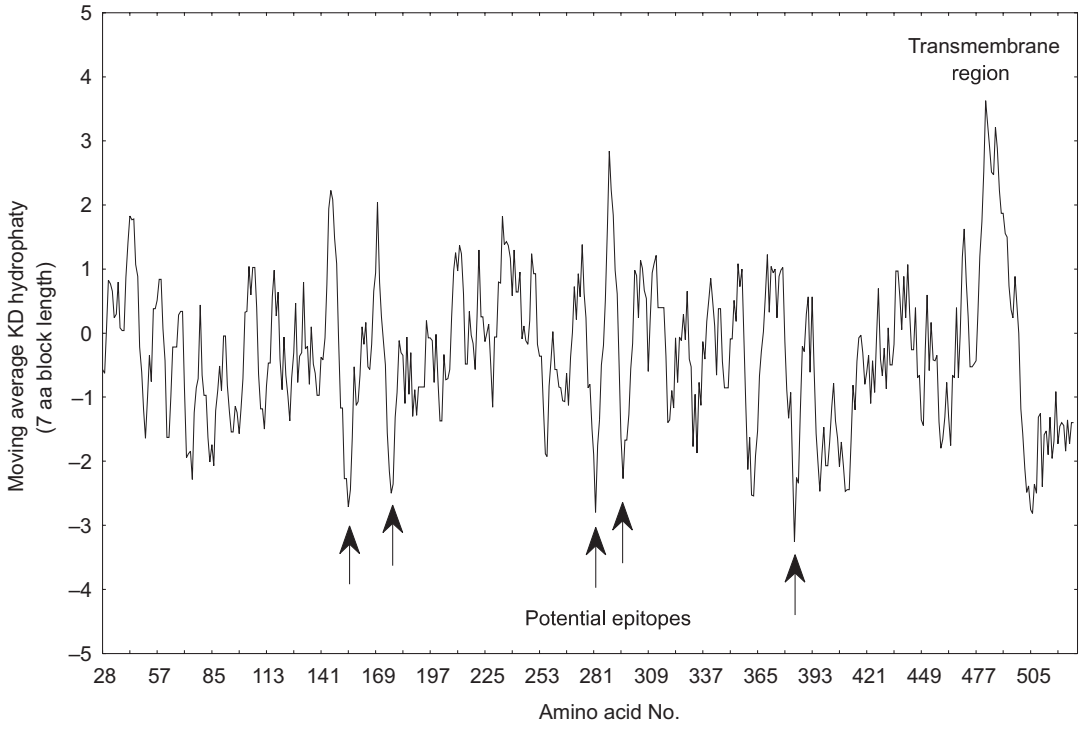


Figure 1. Detection of the potential epitopes within ICAM-1 molecule by means of the Kyte & Doolittle algorithm.

no acids, *i.e.* the signal part, of the ICAM-1 human protein sequence was excluded from the analysis and only the chain part (amino acids 28–532) was used.[3,11–13] Kyte-Doolittle scale is widely applied for determining the hydrophobic character of a protein. This simple, practical and accurate method uses amino acid hydrophaty along the protein chain to predict the positions of exposed and buried residues, antigenic sites, membrane-spanning regions and turns between elements of the secondary structure.[3,11,12] The window size of this scale is the number of amino acids examined at a time to determine a point of hydrophobic character.[3,11,12] The values within particular window are averaged, and a plot is constructed using classical statistical method of the moving average.[3,11] The averaged values are plotted above the central amino acid of the window – therefore the odd number of amino acids is used for the window length.

Window size can be varied from 5 to 25, with the default of 7. The researcher should choose a window corresponding to the expected size of the structural motif that is under investigation. Short window sizes of 5–7 generally work well for predicting the exposed loops/cell epitopes, providing that average values are less than –1.6.[3,11,12] Large window sizes of 19–21 are used for finding the transmembrane domains if the values calculated are above 1.6. [3,11,12] Regions with the values between 0 and 1.6 are defined as hydrophobic, but not transmembrane.

Amino acid hydropathy values averaged over a segment of 7 residues were plotted to predict exposed loops/epitopes. Predicted locations of the 5 epitope sites are presented in Figure 1, as follows: 1. RGEK (aa 152–155), 2. VLV (aa 173–175), 3. EDE (aa 281–283), 4. LGN (294–296) and 5. HKN (383–385).

The first two predicted epitopes are located in the β-strand regions of the molecule (aa 146–152, aa 155–161, and aa 167–174). The position of the first epitope at aa152–154 is an important molecular structure, a cell attachement site.[13] The second one (173–175) is within extracellular repetitive region that may be also suitable for the molecular recognition, and the E residue (aa 281) of the third epitope is important for integrin binding.[13,14] The fourth and the fifth predicted epitope of ICAM-1 are N-linked glycosylation sites at the positions 296 and 385. Transmembrane segment located at the position 481–504 is accurately predicted (Figure 1 – peaks at 481–490). Presented results of the ICAM-1 analyses by means of the Kyte & Doolittle algorithm in Figure 1 confirm the validity, accuracy and applicability of the procedure.

### *Ligand-Receptor Recognition by MRT*

Sense-antisense pairs of amino acides are derived from the genetic code patterns. They are obtained from the mRNA codon sequence transcribed in either 3'→5' (left to right) or 5'→3' (right to left) direction. During this process uracil (U) is transcribed into adenine (A) and cytosine (C) is transcribed into guanine (G), or vice versa.[4–10] Amino acid pairs arising from this genetic code feature are given in Table I.

TABLE I. Sense-antisense pairs of amino acides derived from the genetic code

| Amino acid (codon*) | KD hydropathy | Antisense (3'→5') | Antisense (5'→3') |
|---|---|---|---|
| I (AUU, AUC, AUA) | 4.5 (hydrophobic) | Y | Y, N, D |
| V (GUU, GUC, GUA, GUG) | 4.2 (hydrophobic) | H, Q | H, N, D, Y |
| L (UUA, UUG, CUU, CUC, CUA, CUG) | 3.7 (hydrophobic) | N, E, D | E, K, Q |
| F (UUU, UUC) | 2.7 (hydrophobic) | K | K, E |
| C (UGU, UGC) | 2.5 (hydrophobic) | T | T, A |
| M (AUG) | 1.9 (hydrophobic) | Y | H |
| A (GCU, GCC, GCA, GCG) | 1.8 (hydrophobic) | R | R, G, S, C |
| G (GGU, GGC GGA, GGG) | –0.4 (neutral) | P | P, S, T, A |
| T (ACU, ACC, ACA, ACG) | –0.7 (neutral) | C, W | G, S, C, R |
| W (UGG) | –0.9 (neutral) | T | P |
| S (UCU, UCC, UCG, UCA, AGU, AGC) | –0.9 (neutral) | S, R | R, G, T, A |
| Y (UAU, UAC) | –1.3 (neutral) | M, I | I, V |
| P (CCU, CCC, CCA, CCG) | –1.6 (neutral) | G | G, W, R |
| H (CAU, CAC) | –3.2 (hydrophilic) | V | V, M |
| D (GAU, GAC) | –3.5 (hydrophilic) | L | I, V |
| E (GAA, GAG) | –3.5 (hydrophilic) | L | L, F |
| N (AAU, AAC) | –3.5 (hydrophilic) | L | I, V |
| Q (CAA, CAG) | –3.5 (hydrophilic) | V | L |
| K (AAA, AAG) | –3.9 (hydrophilic) | F | F, L |
| R (CGU, CGC, CGA, CGG, AGA, AGG) | –4.5 (hydrophilic) | A, S | A, S, P, T |

*UAA, UAG and UGA are stop codons

Sense and antisense peptides have mutually complementary shapes which results in their interaction.[4–10] This is due to the fact that their hydrophilic and hydrophobic patterns of amino acid polarity are in most cases opposite,[4–10] a fact clearly shown in Table I.

We observed the relationship of ICAM-1 epitopes and its possible ligands (antisense peptides) that arise from the mRNA sequence transcription in 3'→5' and 5'→3' directions. Leukocyte function-associated antigen-1 (LFA-1) is a natural ligand for the ICAM-1.[3] Consequently, MRT was used to locate possible binding sites for ICAM-1 epitopes. Table II confirms the applicability of MRT for such a purpose since 5 out of 10 possible ligands to ICAM-1 epitopes were found to be located at different positions of the LFA-1 molecule. A ProteinInfo sequence search of the NCBInr database was used to locate ligands derived by means of MRT.[15]

Predicted interacting pairs located at the positions 281–283, 294–296 and 383–385 were not further investigated. Motifs 294–296 and 383–385 overlap with N-linked glycosylation sites (at positions 296 and 385). It is well known that a »cloud« of sugar found at the glycosylation sites influences that paratope could not bind to the exposed epitope.[2]

Two motifs that were further analysed were the first two predicted epitopes in the β-strand regions of the ICAM-1 molecule (positions 152–155 and 173–175, Figure 1). Both of them represent important molecular structures, a cell attachement site and a part of extracellular repetitive region.[13,14] β-strand regions are considered to be more suitable for the MRT based prediction of epitope-paratope interaction since such structural pattern could facilitate binding.[7]

TABLE II. ICAM-1 epitopes according to the Kyte & Doolittle algorithm and predicted LFA-1 ligands defined by MRT

| ICAM-1 | Sense aa | LFA-1 | Antisense aa |
|---|---|---|---|
| 152–155 | RGEK | 482–485 | APLF |
| | | not located | TPLL |
| 173–175 | VLV | 962–964 | HDH |
| | | not located | HQH |
| 281–283 | EDE | 1161–1163 | LLL |
| | | 675–678 | LQL |
| 294–296 | LGN | not located | DPL |
| | | not located | QPV |
| 383–385 | HKN | 158–160 | VFL |
| | | not located | VLV |

Analyses of two relevant ICAM-1 epitope regions (aa 146–155, aa 167–175) and related LFA-1 binding structures were done with SSpro 2.0 and CONpro structure prediction methods (http://www.igb.uci.edu/tools/scratch/). SSpro is a server for protein secondary structure predic-

tion based on an ensemble of 1D-RNNs (one dimensional recurrent neural networks, *i.e*. bidirectional recurrent neural networks).[15–17] Experiments on an independent test set show a performance exceeding 78 % correctly classified residues on the CASP-like assignment of the secondary structure into 3 classes denoted as H (helix), E (strand) and C (the rest). More lenient assignments lead to 80 % or better.[16,17] It is worth mentioning that peptides/proteins shorter than 30 amino acid residues often do not have a well defined structure.[18] The chosen length of the sliding block may also influence the prediction result.[18] SSpro method uses PSI-BLAST to include all homologous proteins in the analysis. Very high levels of local homology to already known structures is used to improve the prediction accuracy and minimaze the prediction error arising from the length of the sequence that is used,[19] which makes SSpro applicable for the analysis of short linear segments within proteins.

CONpro method predicts if the contact of the residue in a protein is above or below the average. The method predicts number of residue contacts at the threshold radius of 12 Å relative to the amino acids average number of contacts.[20,21] It defines if the contact of the residue in a protein is above (+) or below the average (–).[20,21] The prediction is based on an ensemble of 10 artificial neural networks (1D-RNNs), adopting as input a multiple alignment of homologues generated by PSI-BLAST.[20,21] The accuracy of the prediction is 73 %. The overwhealing majority of protein contacts and the best results of CONpro prediction are found at linear distances shorter than 100 amino acids, and the method is relevant for proteins up to a sequence length of 300 amino acids.[22]

Table III presents the prediction results for the epitopes of the ICAM-1 β-strand regions 146–155 and 167–175 containing motifs RGEK at positions 152–155 and VLV at positions 173–175, respectively (Figure 1, Table II). RGE epitope motif is similar to the common integrin-binding RGD sequence.[23] Extended β-strand regions of the sense peptides and contact maps that should facilitate binding (denoted by +) are clearly visible.

TABLE III. Prediction results for the epitopes of the ICAM-1 amino acid regions 146–155 and 167–175

| **sense** | **antisense** |
|---|---|
| L T V V L L R G E K | E W H H D E A P L F |
| C E E E E E C C C C | C C C C C C C C C C |
| – – – + + – – – – – | – – – – – – – – – – |
| **sense** | **antisense** |
| A E V T T T V L V | R L Q C W C H D H |
| C C E E E E E E C | C C C C C C C C C |
| – – – – – – – – – | – – – – + – – – – |

## CONCLUSION

Our analyses based on the well known model of ICAM-1 and LFA-1 molecular interaction indicates that MRT represents an applicable heuristic algorithm for the ligand-receptor interactions. This algorithm achieves better efficacy when being combined to several other molecular modelling procedures, *e.g.* molecular hydropathy analyses, secondary structure prediction methods and protein database search of other relevant structural and functional data. The limitation of the combined MRT-hydropathy analyses is in the fact that it cannot explain 3D protein interactions, but it can be a valuable starting point for a more complex computational and experimental analyses.

*Acknowledgement*. – This work was supported by the Ministry of Science, Education and Sports, Republic of Croatia (Grant No. 098-0982929-2524).

## REFERENCES

1. W. J. Herbert, P. C. Wilkinson, and D. I. Scott, *The Dictionary of Immunology*, Academic Press, San Diego, 1995, pp. 58–59.
2. F. Brown, G. Dougan, E. M. Hoey, S. J. Martin, B. K. Rima, and A. Trudgett, *Vaccine Design,* John Wiley & Sons, Chichester, pp. 33–52.
3. P. A. Bates and M. J. E. Sternberg, *From Protein Sequence to Structure*, in: A. R. Rees, M. J. E. Sternberg, and R. Wetzel (Eds.), *Protein Engineering,* Oxford University Press, Oxford, 1992, pp. 117–141.
4. J. E. Blalock and K. L. Bost, *Biochem. J.* **234** (1986) 679–683.
5. J. E. Blalock, *Nature Medicine* **1** (1995) 876–878.
6. N. Štambuk, *Croat. Chem. Acta* **71** (1998) 573–589.
7. J. R. Heal, G. V. Roberts, J. G. Raynes, A. Bhakoo, and A. D. Miller, *ChemBioChem.* **3** (2002) 136–151.
8. N. Štambuk, P. Konjevoda, A. Boban-Blagaić, and B. Pokrić, *Theory Biosci.* **123** (2005) 265–275.
9. J. R. Heal, G. W. Roberts, G. Christie, and A. D. Miller, *ChemBioChem.* **3** (2002) 86–92.
10. V. J. Hurby and T. O. Matsunaga, *Applications of Synthetic Peptides*, in: G. A. Grant (Ed.), *Synthetic Peptides,* Oxford University Press, Oxford, 2002, pp. 292–376.
11. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157** (1982) 105–132.
12. A. M. Lesk, *Introduction to Bioinformatics*, Oxford University Press, Oxford, 2003, pp. 217–276.
13. Protein ICAM1_HUMAN, Primary accession number P05362, UniProt database, Swiss-Prot entry, Release 47, 10-MAY-2005, (http://beta.uniprot.org/uniprot/P05362).
14. Y. Yang, C. D. Jun, J. H. Liu, R. Zhang, A. Joachimiak, T. A. Springer, and J. H. Wang. *Mol. Cell* **14** (2004) 269–276.
15. K. Houra, N. Štambuk, P. Konjevoda, A. Boban-Blagaić, T. Bruketa, and B. Pokrić. *Croat. Chem. Acta* **79** (2006) 379–383
16. G. Pollastri, D. Przybylski, B. Rost, and P. Baldi, *Proteins* **47** (2002) 228–235.

17. P. Baldi, S. Brunak, P. Frasconi, G. Pollastri, and G. Soda, *Bioinformatics* **15** (1999) 937–946.
18. J. A. Cuff and G. J. Barton, *Proteins: Structure, Function and Genetics* **34** (1999) 508–519.
19. J. Cheng, A. Z. Randall, M. J. Sweredoski, and P. Baldi, *Nucl. Acids. Res*. **33** (2005) W72–W76.
20. G. Pollastri, P. Baldi, P. Fariselli, and R. Casadio, *Bioinformatics* **17** (**Suppl 1**) (2001) S234–S242.
21. G. Pollastri, P. Baldi, P. Fariselli, and R. Casadio, *Proteins* **47** (2002) 142–153.
22. P. Baldi and G. Pollastri, *J. Mach. Learn. Res*, **4** (2003) 575–602.
23. E. Ruoslahti, *Annu. Rev. Cell Dev. Biol*. **12** (1996) 697–715.

## SAŽETAK

### Modeliranje ICAM-1 i LFA-1 interakcije korištenjem teorije molekularnog prepoznavanja

**Nikola Štambuk, Paško Konjevoda, Dražen Vikić-Topić i Biserka Pokrić**

Korišten je model interakcije ICAM-1 i LFA-1 molekula kako bi se provjerila primjena teorije molekularnog prepoznavanja u identifikaciji različitih veznih mjesta, to jest receptorskih veznih mjesta unutar većih antigenskih molekula. Rezultat ukazuje kako je teorija molekularnog prepoznavanja heuristički algoritam uporabljiv za identifikaciju epitopa/paratopa unutar većih molekula. Međutim, kako bi se postigla veća učinkovitost, ovaj se heuristički algoritam molekularnog prepoznavanja treba kombinirati s više drugih postupaka: analiza molekularnih hidropatija, postupci za predikciju sekundarne strukture i pretraživanje proteinskih baza podataka. Ograničenje kombinirane analize pomoću teorije molekularnog prepoznavanja i hidropatija jest u činjenici da se time ne mogu objasniti 3D proteinske interakcije, ali može poslužiti kao korisna početna točka za kompleksnije izračunske i eksperimentalne analize.