Putative hAPN receptor binding sites in SARS_CoV spike protein[1]

YU Xiao-Jing[2], LUO Cheng[3], LIN Jian-Cheng[2], HAO Pei[2], HE You-Yu[2], GUO Zong-Ming[4], QIN Lei[4], SU Jiong[4], LIU Bo-Shu[2], HUANG Yin[2], NAN Peng[4,5], LI Chuan-Song[2,6], XIONG Bin[3], LUO Xiao-Min[3], ZHAO Guo-Ping[2], PEI Gang[2], CHEN Kai-Xian[2], SHEN Xu[3*], SHEN Jian-Hua[3*], ZOU Jian-Ping[3*], HE Wei-Zhong[4*], SHI Tie-Liu[2*], ZHONG Yang[4,5*], JIANG Hua-Liang[3], LI Yi-Xue[2,4*]

[2]Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031;
[3]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201203;
[4]Shanghai Center for Bioinformation Technology, Shanghai 201203; [5]School of Life Sciences, Fudan University, Shanghai 200433;
[6]Institute of Medical Biology, Chinese Academy of Medical Science, Yunnan 650118, China

KEY WORDS severe acute respiratory syndrome (SARS); CD13; spike protein; bioinformatics; molecular modeling; molecular docking

ABSTRACT

AIM: To obtain the information of ligand-receptor binding between the S protein of SARS_CoV and CD13, identify the possible interacting domains or motifs related to binding sites, and provide clues for studying the functions of SARS proteins and designing anti-SARS drugs and vaccines. **METHODS:** On the basis of comparative genomics, the homology search, phylogenetic analyses, and multi-sequence alignment were used to predict CD13 related interacting domains and binding sites in the S protein of SARS_CoV. Molecular modeling and docking simulation methods were employed to address the interaction feature between CD13 and S protein of SARS_CoV in validating the bioinformatics predictions. **RESULTS:** Possible binding sites in the SARS_CoV S protein to CD13 have been mapped out by using bioinformatics analysis tools. The binding for one protein-protein interaction pair (D757-R761 motif of the SARS_CoV S protein to P585-A653 domain of CD13) has been simulated by molecular modeling and docking simulation methods. **CONCLUSION:** CD13 may be a possible receptor of the SARS_CoV S protein, which may be associated with the SARS infection. This study also provides a possible strategy for mapping the possible binding receptors of the proteins in a genome.

[1] Project supported by the 863 Hi-Tech Program, (No 2001AA235051, 2001AA235071, 2001AA233031), the National Natural Science Foundation of China, (No 29725203, 20072042), the State Key Program of Basic Research of China, (No 2002CB512801, 2002CB512802, 2002CB512801, 110CB510209), the special programs of oppugning SARS from the Ministry of Science and Technology, Chinese Academy of Sciences and the Shanghai Science and Technology Commission, Key Program of Basic Research of Science & Technology Commission of Shanghai Municipality, (No 02DJ14013, 02DJ14014), Shanghai Basic Research Project from the Shanghai Science and Technology Commission, (No 03DZ14102, 02DJ14070), and the One Hundred People Program of Chinese Academy of Sciences (1999), and the Knowledge Innovation Program of Chinese Academy of Sciences, (No KSCX2-SW-202).
[*] Correspondence to Prof LI Yi-Xue, SHI Tie-Liu, ZHONG Yang, HE Wei-Zhong, JIANG Hua-Liang, SHEN Jian-Hua, SHEN Xu, and ZOU Jian-Ping.
Received 2003-05-13 Accepted 2003-05-15

INTRODUCTION

The genome of severe acute respiratory syndrome associated with a special coronaviruses (SARS_CoV) was sequenced just two weeks ago and is now available in the GenBank[1,2]. It can be found that SARS_CoV genome has the same frame of sequence structure with other coronaviruses, because almost all of the structure proteins existing in previously known coronaviruses, such as spike glycoprotein (S), envelope protein (E), membrane protein (M) and nucleocapsid protein (N), have been identified in SARS_CoV in the same order[3]. However, phylogenetic analyses of the predicted viral proteins indicated that SARS_CoV did not closely resemble any of the three previously known groups of coronaviruses[4,5].

Among the structure proteins of coronaviruses, the S protein plays a very important role in virus entry, virus-receptor interactions and their relationship to tropism. S protein is a surface projection glycoprotein and may be cleaved by virus-encoded or host-encoded proteinases to produce two functional subunits, S1 and S2, like other coronaviruses[3]. Tresnan *et al* reported that two members of coronavirus serogroup I, human respiratory coronavirus (HCoV-229E) and porcine transmissible gastroenteritis virus (TGEV), use aminopeptidase N (APN) as their cellular receptor[6]. The S protein of SARS_CoV is a very large membrane glycoprotein, predicted to be 1 255 amino acids in length. It contains two hydrophobic regions characterized as type 1 glycoproteins: one at the N terminus of the entire protein including a short type I signal sequence and the other at the C terminus consisting of a transmembrane domain and a cytoplasmic tail rich in highly conserved cysteine residues.

It is already known that in infection progresses, a virus can encode different kinds of proteins, which are similar to chemokines or inhibitors of chemokine receptor, simulating the function of chemokine or its receptor to disturb transferring of chemotactic signal, and reducing the immunoreaction mediated by cytotoxic T-lymphocytes (CTL). A lot of researches revealed that the S protein had several important functions including binding of the virus to susceptible cells, mediation of membrane fusion (both viruscell and cell-cell fusions), and induction of neutralizing antibody responses in the host species[7-11]. All of these functions of the S protein are associated with virus infection. This gives us a clue to propose a hypothesis that certain domain(s) in spike protein of SARS_CoV might have the same function as that of HCoV-229E spike glycoproteins–exhibiting its infection function by binding to the human aminopeptidase N (hAPN) (or termed CD13).

Generally, like the S protein, membrane protein receptors are a class of proteins that are pivotal for numerous biological functions. Knowledge of the ligand–receptor binding and activation is critical in understanding the biological functions of receptors and in designing small molecules for therapeutic intervention. Accordingly, to find evidences for our above hypothesis that whether the interactions between the S protein of SARS_CoV and CD13 exist or not, protein-protein interaction analyses were carried out by bioinformatics analysis and molecular modeling methods. In the present paper, we report a putative binding domain and binding sites of the SARS_CoV S protein to CD13.

MATERIALS AND METHODS

Protein-protein interaction mapping Two kinds of sequence alignments were initially performed: the first one was carried out between the S protein of SARS_CoV (GenBank protein ID NP_828851.1, Genome ID NC_004718, SARS coronavirus TOR2, submitted by the British Columbia Centre for Disease Control and National Microbiology Laboratory of Canada) and the S protein of HCoV-229E; the second one was undertaken between the S protein of SARS_CoV (GenBank protein ID NP_828851.1, Genome ID NC_004718, SARS coronavirus TOR2) and the S protein of porcine epidemic diarrhea virus (PEDV). The first alignment produced 29 % identities, 45 % positives and 13 % gaps, respectively, and the second alignment revealed 29 % identities, 45 % positives and 12 % gaps, respectively. These two alignments addressed the region with the highest similarity for three of the S proteins located at the area around the S2 segment, but there is no significant homology among them. Similar results of the multi-alignment analysis were obtained by using CLUSTAL_W. These results are in consistent with the current phylogenetic analyses of Rota *et al*[1], Marra *et al*[2], Thomas *et al*[4], and Peiris *et al*[5]. Based on the results of multi-alignment and phylogenetic analyses, they suggested that SARS_CoV should be sorted as a new group of coronaviruses. This means that only through sequence alignments employing available bioinformatics analysis tools such as BLAST and CLUSTAL_W cannot achieve enough useful information for the ligand-receptor interactions between the S protein of SARS_CoV and the CD13.

Because of aforementioned reason, a new developed method of our laboratory was applied to map the binding sites of the S protein of SARS_CoV to CD13. In brief, the procedure of this method composed following steps: (1) Homology search was performed to identify the homologous proteins related to CD13. (2) A phylogenetic tree was constructed to determine proteins that are homologous and have short phylogenetic distances to CD13 (Fig 1), these proteins were selected as queries for searching the protein-protein interaction database DIP (http://www.DIP.org). (3) Proteins in the DIP database that possibly interact with CD13 and their homologous proteins with short phylogenetic distances to CD13 were picked out to construct a database, designated "InterDatabase". InterDatabase contains 11 proteins: PEP3_YEAST (GenBank ID P27801), BZZ1_YEAST BZZ1 (GenBank ID P38822), PRY3_YEAST PRY3 (GenBank ID P47033), PR06_YEAST (GenBank ID P19735), NUT2_YEAST NUT2 (GenBank ID Q06213), GAS1_YEAST (GenBank ID P22146), MUCB_ECOLI (GenBank ID P07375), MUCA_SALTY (GenBank ID P07376), RECA_ECOLI (GenBank ID P03017), SSB_ECOLI (GenBank ID P02339), YTM1_YEAST (GenBank ID Q12024). Practically, proteins in InterDatabase can be taken as the templates for searching ligands of CD13. (4) Taking all the putative proteins of SARS_CoV as queries, InterDatabase was searched by using BlastP and PSI-Blast, addressing the proteins in SARS_CoV genome that are mostly homologous to the proteins in the InterDatabase. These proteins can be taken as the ligand candidates of CD13. (5) Finally, a second database was constructed containing the protein interaction pairs between CD13 and releted proteins and proteins in the SARS_CoV genome, names as CD13-SARS-PPI, for further analysis and study.

Molecular modeling To confirm above protein-protein interactions, the possible three-dimensional (3D) structures of the monomer proteins in CD13-SARS-PPI database were constructed by molecular modeling methods, and the protein-protein interactions were simulated by molecular docking and molecular dynamics methods. In the present paper, we just report the simulation result for one candidate of the interaction between CD13 and SARS_CoV proteins.

3D model generation of CD13 Sequence analy-

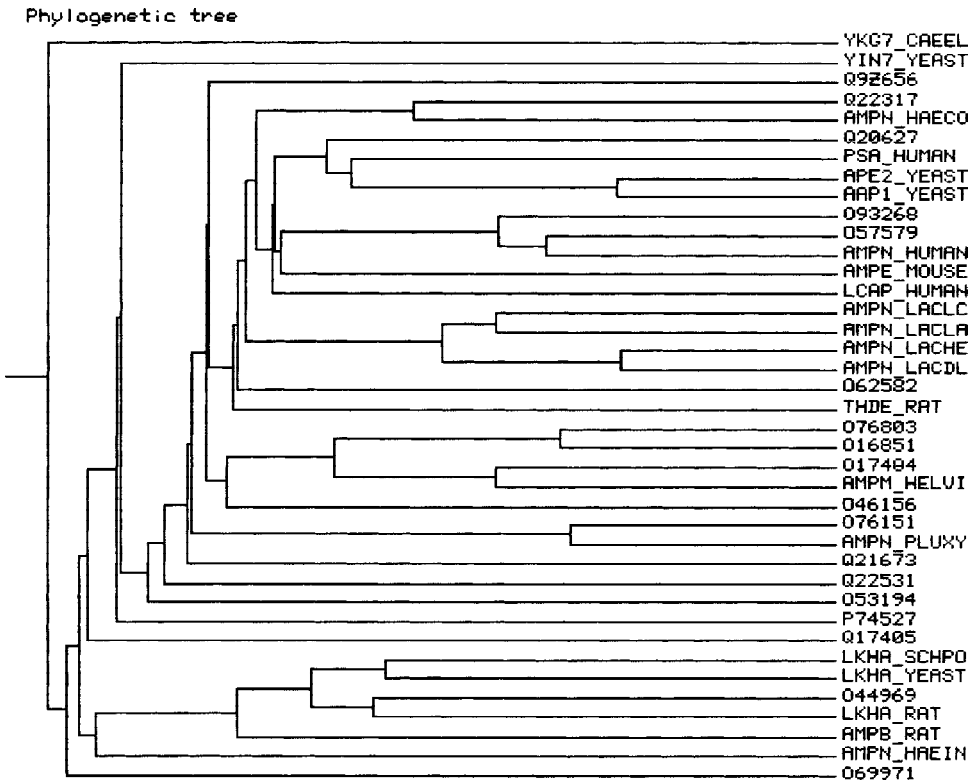

Fig 1. A phylogenetic tree based on the proteins which are homologous and have short phylogenetic distances to CD13. These proteins were selected as queries to search protein-protein interaction database DIP (http://www.DIP.org) to find CD13 related information of protein-protein interactions. Eleven proteins were selected to build up an InterDatabase.

sis and conserved-residue identification were carried out among CD13 and other aminopetidases, sequences of CD13 and other aminopetidases were from (http://www.expasy.ch/prosite). Protein-protein interaction analyses indicate that sequence from Asp757 to Arg761 (termed D757-R761) in the SARS_CoV S protein is a possible binding motif to CD13, and its corresponding binding domain in CD13 is around the segment from Pro585 to Ala653 (termed P585-A653). It is impossible to model the whole structures of the full length CD13 containing 966 amino acids (Swiss-Prot ID P15144). Therefore, we selected sequence from Asp438 to Phe814 of CD13 (termed D438-P814), which covers P585-A653 domain, for structure modeling. The crystal structure of leukotrinen A4 hydrolase/aminopeptidase[12] was selected as the template for homology modeling. During the 3D structure construction, the Homology module encoded in Insight II[13] and the ClustalW algorithm[14] were applied in sequence alignment, and the Blosum scoring matrix[15] was employed to obtain the best-fit alignment. The best alignment was selected according to not only the value of the alignment score but also the reciprocal positions of conserved residues. The transmembrane (TM) domains were identified and transformed into α-helices and the nonequivalent amino acids were mutated to produce CD13 sequence.

The FASTA program[16] was used to identify sequence homologous through the in-house database[17] containing 700 loops and proteins with medium to high sequence identity. CLUSTAL_W[14] was then used to determine the fragments that have higher homology with the loops and the N terminus of the CD13. The reasonable fragment conformation was chosen from the top 10 candidates that have the lowest root mean square (RMS) values and considerable geometrical compatibility.

The 3D model of D438-P814 was optimized by molecular mechanics method with Amber force field and Kollman-all-atom charges[18]. Finally, the HBPLUS algorithm[19] was used to calculate the inter-helical hydrogen bonds, and the modeled structure was validated with PROCHECK[20] and WHATIF[21].

Interaction simulation of D757-R761 to D438-P814 To address the interaction feature between the key motif of SARS_CoV S protein (D757-R761) and its possible binding site in CD13 (P585-A653), molecular docking simulation was performed. Because the polypeptide of D757-R761 (Asp757-Arg758-Asn759-Thr760-Arg761) is very flexible, a powerful computational searching method is needed. Genetic algorithm-based flexible docking program, FlexiDock[22], provides a means for the simulation of flexible ligand binding to its receptor. FlexiDock incorporates the van der Waals, electrostatic, torsional and constraint energy terms of the Tripos force field. And it uses a genetic algorithm to determine the optimum ligand geometry. The structures of D438-P814 of CD13 and D757-R761 of SARS_CoV S protein were prepared using the Sybyl 6.8[23] molecular modeling software. After FlexiDock, we gained 20 possible binding conformations for D757-R761, and the conformation with the lowest binding energy to D438-P814 was selected for further analyses.

RESULTS AND DISCUSSION

Possible interaction pairs between SARS_CoV and CD13 Proteins with E-value<1 estimated by our new developed method were selected as the candidates for protein-protein interaction. Three protein candidates in SARS_CoV genome that may interact with CD13 have been identified by this method, they are (1) the S protein with five BLAST hits regions, BLAST score from 29.0 to 23.1, 273 bits and the best E-value of 0.015; (2) coronavirus nsp5 protein (GenBank ID NP_828866.1), with one BLAST hits regions, 52 hits, BLAST score of 24.7 and the best E-value of 0.039; and (3) putative uncharacterized protein (GenBank ID NP_828857.1) with one BLAST hits regions, 55 hits, BLAST score of 25.8 and the best E-value of 0.041. Our bioinformatics analysis strongly suggested that domains and motifs existed in the S protein of SARS_CoV that may interact with CD13, but other two protein candidates might result from false positive prediction because of their lower hits and higher E-values.

To confirm above result, we performed a multi-alignment for the sequences among the S proteins of SARS_CoV, HCoV-229E, and PEDV as well as other proteins in the InterDatabase. An extension homologous domain (aa. 503-1187) and a common homologous domain (aa. 1132-1171) were identified from the S proteins of SARS_CoV. The common homologous domain contains 49 residues (EELDKYFKNHTSPDV-DLGDISGINASVVNIQKEIDRLNEV), located close to the C-terminus, 24 amino acids far away from the predicted transmembrane domain of SARS_CoV S protein[2]. Two binding sites for CD13 were detected at the segment of aa. 503-1187 in the SARS_CoV S protein,

ie, TLVK (aa. 943-946) and FVF (aa. 1077-1079), which respectively correspond to the two CD13 binding sites of TLLL (aa. 62-65) and YAY (aa. 214-216) in the S protein of MHV[8]. Interestingly, the number of amino acids (~131 aa.) between the two detected binding sites are approximately as same as that between the two binding sites of the S protein of MHV (~151 aa.) (Fig 2). Multi-alignment for S proteins of seven coronaviruses resulted in four conserved motifs (Fig 3), three of them have IPTNFSIS (aa. 695-703),

```
seq 1. 932 ---TLVK----FVF---- 1090,970 ---TXXL--- 993
           **+     + +          * *
seq 2.  51 ---TLLL----YAY----  226, 51 ---TLLL---  75
```

Fig 2. Three putative binding sites for CD13 in SARS_CoV S protein, TLVX (aa. 943-946), TXXL (aa. 980-983) and FVF (aa. 1077-1079) (seq_1), which are in agreement with CD13 binding motifs of MHVR in the same order (seq_2).

RSFIEDLLFNK (aa. 797-808), and DRLITGRLQSLQ (aa. 976-987), located at the extension region 503-1187. A possible binding site TXXL (aa. 980-983) corresponding to TLLL of MHVR was also captured (Fig 2).

Additionally, the whole extension homologous domain of SARS_CoV S protein was scanned by a physical chemistry property-based receptor-binding domain searching method[24-26] in order to find some possible extra CD13 binding sites. One possible CD13 binding motif in the S protein of SARS_CoV was identified, viz DRNTR (D757-R761). The whole sequence of CD13 was also scanned by the same method, resulting the possible binding domain of CD13 to D757-R761, ie, PITSIRDGRQQQDYWLIDVRAQNDLFSTSGNEWVLLNLNVTGYYRVNYDEENWRKIQTQLQRDHSA (P585-A653). Experimental research reported that a CD13 binding domain ranging from amino acids 417 to 547 existed in the S protein of HCoV-229E[11]. Sequence

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                 660                                                710
HCoV-SARS        YTMSLGADSS IAYSNN---- -----TIAIP TNFSISITTE VMPVSMAKTS
HCoV-229E        PRNVSYDSVS AIVTAN---- ------LSIP SNWTTSVQVE YLQITSTPIV
Porcine_tr       VTHSDGD-VQ PISTGN---- ------VTIP TNFTISVQVE YIQVYTTPVS
Porcine_re       VTHSDGD-VQ PISTGN---- ------VTIP TNFTISVQVE YIQVYTTPVS
Murine sp        YRLTTFEPYT PMLVNDSVQS VDGLYEMQIP TNFTIGHHEE FIQTRSPKVT
Bovine sp        YRFTNFEPFT VNSVNDSLEP VGGLYEIQIP SEFTIGNMEE FIQISSPKVT
Canine           VTHSDGD-VQ PISTGN---- ------VTIP TNFTISVQVE YIQVYTTPVS
Consensus                     N---- -----  IP TNFTIS  E  IQ    V
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                 790                                                840
HCoV-SARS        ------RSFI EDLLFNKVTL ADAG-FMKQY GECLGDINAR DLICAQKFNG
HCoV-229E        ------RSAI EDILFSKLVT SGLGTVDADY KKCTKGLSIA DLACAQYYNG
Porcine_tr       ------RSAI EDLLFSKVVT SGLGTVDEDY KRCTGGYDIA DLVCAQYYNG
Porcine_re       ------RSAI EDLLFSKVVT SGLGTVDEDY KRCTGGYDIA DLVCAQYYNG
Murine sp        PSAIRGRSAI EDLLFDKVKL SDVG-FVEAY NNCTGGQEVR DLLCVQSFNG
Bovine sp        ------RSAI EDLLFSKVKL SDVG-FVEAY NNCTGGAEIR DLICVQSYNG
Canine           ------RSAI EDLLFDKVVT SGLGTVDEDY KRSAGGYDIA DLVCARYYNG
Consensus        ------RSAI EDLLFSKV   S  G   E Y   CTGG I  DL CAQ YNG
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                 976                                                1026
HCoV-SARS        DRLITGRLQS LQTYVTQQLI RAAEIRASAN LAATKMSECV LGQSKRVDFC
HCoV-229E        DRLITGRLAA LNVFVSHTLT KYTEVRASRQ LAQQKVNECV KSQSKRYGFC
Porcine_tr       DRLITGRLTA LNAFVSQTLT RQAEVRASRQ LAKDKVNECV RSQSQRFGFC
Porcine_re       DRLITGRLTA LNAFVSQTLT RQAEVRASRQ LAKDKVNECV KSQSHRFGFC
Murine sp        DRLINGRLTA LNAYISKQLS DSTLIKVSAA QAIEKVNECV KSQTTRINFC
Bovine sp        DRLINGRLTA LNAYVSQQLS DSTLVKFSAA QAMEKVNECV KSQSSRINFC
Canine           DRLITGRLTA LNAFVSQTLT RQAEVRASRQ LAKDKVNECV RSQSQRFGFC
Consensus        DRLITGRLTA LNA VSQ L     EVRAS  LA  KVNECV  SQS R  FC
```

Fig 3. Multi-sequence alignment among spike proteins of seven coronavirus. Four important motifs that may act as the binding sites to CD13 are IPTNFSIS (aa. 695-703), RSFIEDLLFNK (aa. 797-808), DRLITGRLQSLQ (aa. 976-987), and TXXL.

alignment of the SARS_CoV S protein to HCoV-229E S protein indicated that the CD13 binding domain for the HCoV-229E S protein corresponded to the domain of aa. 500-810, which covered the possible CD13 binding motifs selected out by above method. This demonstrates that our searching method is reliable and that the CD13 binding motifs are reasonable.

All above possible protein-protein interaction informations were deposited in the CD13-SARS-PPI database.

Binding for D438-P814 of CD13 and D757-R761 of SARS_CoV S protein To confirm the sequence analysis result for protein-protein interaction at the 3D level, we are modeling the 3D structures of the monomer proteins and their interactions in the CD13-SARS-PPI database. As an example, in the present paper, we just report the simulation result of the binding between D438-P814 of CD13 and D757-R761 of SARS_CoV S protein.

The 3D structure model of D438-P814 is shown in Fig 4A. PROCHECK[20] and WHATIF[21] analyses indicate that the 3D structure is reasonable. In agreement with above protein-protein interaction prediction, there is a cleft around the possible CD13 binding domain (P585-A653) (Fig 4A).

The binding model derived from the FlexiDock simulation is shown in Fig 4B, which indicates that D757-R761 may complementally fit into the binding cleft on the surface of D438-P814. The binding affinity was estimated about at the level of mmol/L. This illustrates that SARS_CoV S protein may possibly bind to CD13. In detail, the carbonyl and NH_2 groups of Arg758 on the S protein form two hydrogen bonding respectively with the backbone NH group of Trp583 and the carbonyl group of Leu683 on the CD13, the hydrogen bond lengths are 3.03 and 2.56 Å, respectively. Hydrophobic interaction plays an important role in the binding of the SARS_CoV S protein and CD13. Tab 1 lists the major hydrophobic interaction pairs between the D757-R761 motif of SARS_CoV S protein and the D438-P814 domain of CD13. In general Phe579, Ile584, Val585, Phe611, Ile686, and Glu688 of the D438-P814 domain are involved in the hydrophobic interaction to the D757-R761 motif of SARS_CoV S protein. These strong hydrogen bonds and hydrophobic interactions dominantly contribute to the binding of the D757-R761 motif of SARS_CoV S protein to CD13. This primary molecular modeling and simulation demonstrated that the bioinformatics analyses for protein-protein interactions made sense, suggesting several clues for the functional study on the SARS proteins.

Tab 1. The hydrophobic contacts between D757-R761 of SARS_CoV S protein and D438-P814 domain of CD13.

D757-R761 of SARS_CoV S protein		D438-P814 of CD13		Distance (Å)
Residue	Atom	Residue	Atom	
Arg761	CB	Phe611	CZ	3.83
Arg761	CB	Phe611	CE1	3.74
Arg761	C	Phe611	CZ	3.65
Arg761	C	Phe611	CE2	3.70
Arg761	C	Phe611	CE1	3.50
Arg761	C	Phe611	CD2	3.65
Arg761	C	Phe611	CD1	3.47
Arg761	C	Phe611	CG	3.57
Arg761	CA	Phe611	CE1	3.71
Thr760	CG2	Leu683	CD2	3.23
Arg761	CG2	Leu683	CD1	3.46
Arg761	CG2	Leu683	CG	3.55
Arg761	CB	Leu683	CD2	3.26
Arg761	CB	Leu683	CD1	3.57
Arg761	CB	Leu683	CG	3.27
Arg761	CB	Leu683	CB	3.51
Arg761	C	Ile584	CD1	3.85
Arg761	CG	Val585	CG2	3.05
Asn759	CB	Ile584	CG1	3.66
Asn759	CB	Ile584	CA	3.73
Asn759	CB	Trp583	C	3.39
Asn759	C	Ile686	CD1	3.70
Arg758	CZ	Glu688	CG	2.86
Arg758	CZ	Glu688	CB	3.55
Arg758	CZ	Glu688	CA	3.32
Arg758	CZ	Leu683	CB	3.89
Arg758	CD	Glu688	CG	3.86
Arg758	CD	Ile686	CG2	3.59
Arg758	CD	Ile686	CB	3.58
Arg758	CD	Ile686	C	3.67
Arg758	CG	Phe579	CD1	3.57
Arg758	CB	Phe579	CD1	3.60
Asp757	CG	Trp583	CD1	3.59
Asp757	CG	Trp583	CG	3.57
Asp757	CG	Trp583	CB	3.37
Asp757	CA	Phe579	CZ	3.82

CONCLUSIONS

Possible binding sites in the SARS_CoV S protein to CD13 have been mapped out by using bioinformatics analysis tools. The binding for one protein-protein interaction pair (D757-R761 motif of the SARS_CoV S

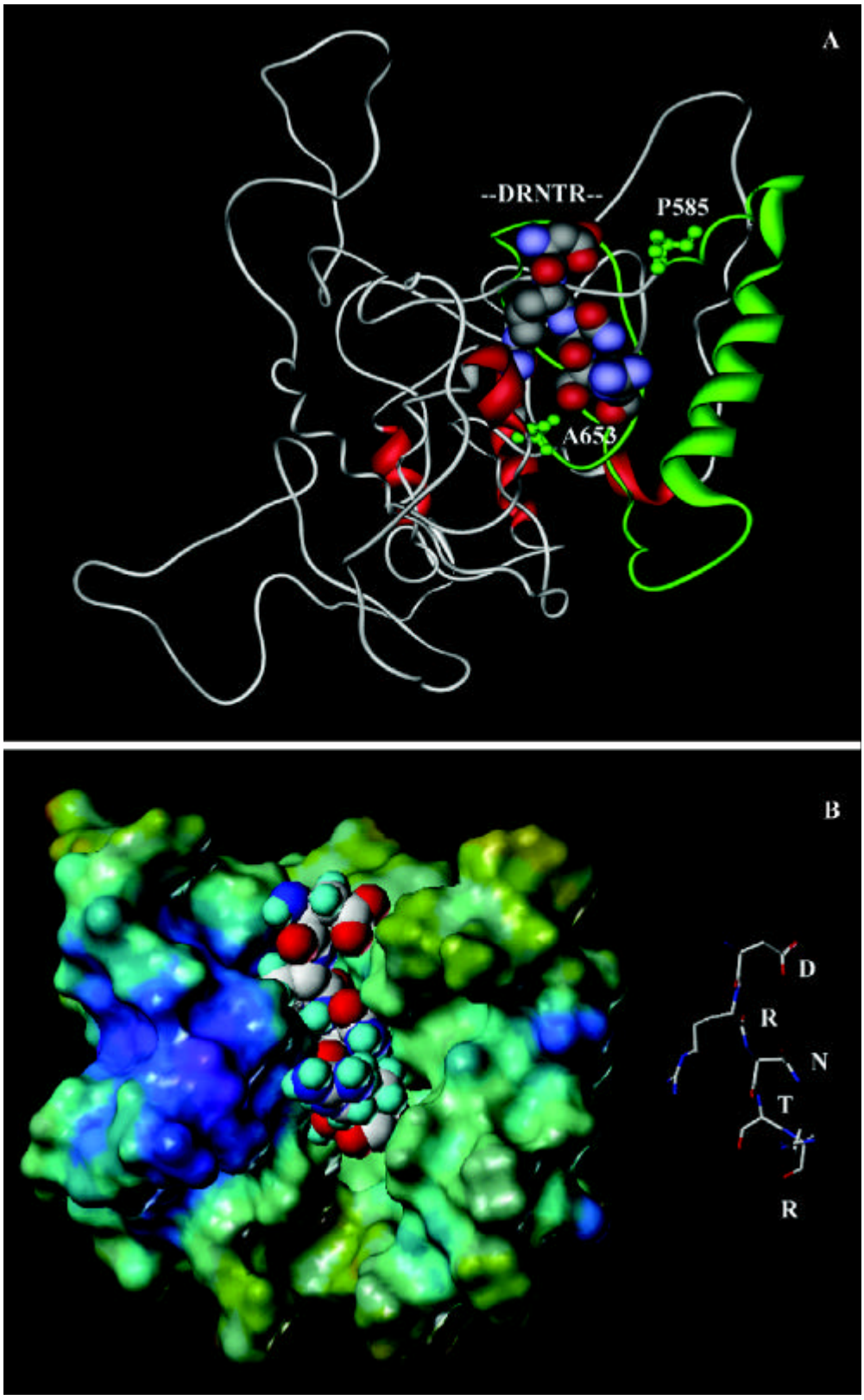

Fig 4. (A) 3D model of D438-P814 of CD13. The structure is shown as ribbon representation. The binding domain of CD13 interacting with the S protein is highlighted in green color, and the starting and end residues (P585 and A653) are represented as ball-and-stick model. (B) The electrostatic surface representation of the interaction between D438-P814 of CD13 and D757-R761 of SARS_CoV S protein. The right part gives the stick representation of the binding conformation of D757-R761.

protein to P585-A653 domain of CD13) has been simulated by molecular modeling and docking simulation methods. The result confirmed the bioinformatics predictions, indicating that CD13 is a possible receptor of the SARS_CoV S protein, which may be associated with the SARS infection. This study also provides a possible strategy for mapping the possible binding sites of the proteins in a genome to their receptors.

REFERENCES

1 Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, *et al*. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science (Sciencexpress). 2003 May 1.

2 Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, *et al.* The genome sequence of the SARS-associated coronavirus. Science (Sciencexpress). 2003 May 1.

3 John Z, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J General Virol 2000; 81: 853–79.

4 Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, *et al*. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003; 1-12.

5 Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LY, Lim W. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 2003; 361: 1319-25.

6 Tresnan DB, Levis R, Holmes KV. Feline Aminopeptidase N (CD13) serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. J Virol 1996; 70: 8669–74.

7 Hingley ST, Leparc-Goffart I, Weiss SR. The spike protein of murine coronavirus mouse hepatitis virus strain A59 is not cleaved in primary glial cells and primary hepatocytes. J Virol 1998; 72: 1606–9.

8 Wessner DR, Shick PC, Lu JH, Cardellichio CB, Gagneten SE, Beauchemin N, *et al*. Mutational analysis of the virus and monoclonal antibody binding sites in MHVR, the cellular receptor of the murine coronavirus mouse hepatitis virus strain A59. J Virol 1998; 72: 1941–8.

9 Arbour N, Ekandé S, Côté G, Lachance C, Chagnon FL. Persistent infection of human oligodendrocytic and neuroglial cell lines by human coronavirus 229E, Am Soc Microbiol 1999; 73: 3326–37.

10 Saeki K, Ohtsuka N, Taguchi F. Identification of spike protein residues of murine coronavirus responsible for receptor-binding activity by use of soluble receptor-resistant mutants. Am Soc Microbiol 1997; 71: 9024–31.

11 Bonavia A, Zelus BD, Wentworth DE, Talbot PJ, Holmes KV. Identiacation of a receptor-binding domain of the spike glycoprotein of human coronavirus HcoVE-229E. J Virol 2003; 77: 2530–8.

12 Rudberg PC, Tholander F, Thunnissen MMGM, Haeggstrom JZ. Leukotriene A4 hydrolase/aminopeptidase, glutamate 271 is a catalyticresidue with specific roles in two distinct enzyme mechanisms. J Biol Chem 2002; 277: 1398.

13 InsightII [molecular modeling package], Version 2000. Calif (USA): Molecular Simulations Inc.

14 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673-80.

15 Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 1992; 89: 10915-9.

16 Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 1990; 183: 63-98.

17 Huang XQ, Jiang, HL, Luo XM, Chen KX, Ji RY, Cao Y, *et al.* Building three-dimensional structures of HIV-1 coreceptor CCR5 and its interaction with antagonist TAK779 by comparative molecular modeling. Acta Pharmacol Sin 2000; 21: 521-8.

18 Cornell WD, Cieplak DP, Bayly CI, Gould IR, Merz KM, Ferguson DM, *et al*. A Second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995; 117: 5179-97.

19 McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994; 238: 777-93.

20 Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993; 26: 283-91.

21 Vriend G, Sander C. Quality control of protein models: directional atomic contact analysis. J Appl Crystallogr 1993; 26: 47-60.

22 Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 1995; 8: 127-34.

23 Sybyl [molecular modeling package], version 6.8. St Louis (MO): Tripos Associates; 2000. p57-1769.

24 Gallet X, Charloteaux B, Thomas A, Brasseur R. A fast method to predict protein interaction sites from sequences. J Mol Biol 2000; 302: 917-26.

25 Eisenberg D, Weiss RM, Terwilliger TC. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 1982; 299: 371-4.

26 Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 1984; 179: 125-42.