that the equilibrium constants are more sensitive to the influence of the substituents in the phenolic ring. In the case of 2,4-dichlorophenol, for example, the formation constants are less compared to other phenols. This may be accounted for by the inductive effect of OH and the steric influence of the Cl atom, which inhibit the complex formation. In nitrophenol–amide systems, K_{11} values are significantly higher due to the greater acidity of the proton donor.

Ratajczak and Orville–Thomas[17] tried to explain the physical origin of such complexes on the basis of Mulliken's charge transfer (C–T) theory, which considers these complexes as a hybrid resonating between a nonpolar structure and a polar one, resulting from the transfer of one electron. Using the C–T theory, we have calculated the fractional charge transfer that may take place in such complexation and the accompanying dipolar increment[18,19] from the shift in υ(O–H) frequencies. Our results indicate that even in the extreme cases where the charge transfer is more or less complete in the ground state, the dipolar increment cannot exceed 0.1 D for these systems. It appears therefore that the C–T theory exaggerates the importance of electron transfer, neglecting other types of interactions.

Many authors have discussed the nature of these interactions on the basis of polarization interactions[20–23]. If one of the two sp^2 hybrid orbitals of the oxygen in the carboxyl group is collinear with the O–H...O axis, the interaction between the polar O–H and the C=O is maximum. Then the O–H bond induces a moment in the highly polarizable lone-pair charge cloud and enhances the IR intensity. Our integrated intensity measurements of the complexed and uncomplexed carbonyl bands showed an increase of 10 to 40% in the bond moment derivatives (Table 3). This shows that polarization interaction is more dominant than electron transfer alone. A similar but enhanced charge in the polarity of the O–H bond is also expected, resulting in a further increase in dipolar increment on complexation.

1. Pimental, G. C. and McClellan, A. L., *The Hydrogen Bond*, W. H. Freeman and Co., Inc., San Francisco, 1960.
2. Gramstad, T. and Funglevik, W. J., *Acta Chem. Scand.*, 1962, **16**, 1369.
3. Schmulbach, C. D. and Hert, D. M., *J. Org. Chem.*, 1964, **29**, 3122.
4. Leroux, C., Samyn, C. and Zeegers-Huyskens, Th., *J. Mol. Struct.*, 1998, **448**, 209.
5. Vanlerheydon, L. and Zeegers-Huyskens, Th., *J. Mol. Liquids*, 1983, **25**, 1.
6. Thijs, R. and Zeegers-Huyskens, Th., *Spectrochim. Acta, Part A*, 1984, **40**, 307.
7. Dorval, C. and Zeegers-Huyskens, Th., *Spectrochim. Acta, Part A*, 1973, **29**, 1805.
8. Whetsel, K. B. and Kagarise, R. E., *Spectrochim. Acta*, 1962, **18**, 315; 329; 341.
9. Sriraman, S., Ramaswamy, K. and Shanmugasundaram, V., *J. Chem. Phys.*, 1967, **64**, 957.
10. Krishnapillai, M. G., Ramaswamy, K. and Gnanadesiken, S. G., *J. Mol. Spectrosc.*, 1965, **17**, 370.
11. Sriraman, S. and Sabesan, R., *J. Chem. Phys.*, 1966, **63**, 716.
12. Freeman, D. E., *J. Opt. Am.*, 1962, **52**, 103.
13. Inuzuka, K., Ito, M. and Imanishi, S., *Bull. Chem. Soc. Jpn.*, 1961, **34**, 476.
14. Sriraman, S., Ramaswamy, K. and Shanmugasundaram, V., *J. Chem. Phys.*, 1967, **64**, 957.
15. Shanmugasundaram, V. and Meyyappan, M., *Acta Chim. Acad. Sci. Hung.*, 1975, **87**, 239.
16. Krishnapillai, M. G., Ramaswamy, K. and Pichai, R., *Indian J. Chem.*, 1965, **3**, 510.
17. Ratajczak, H. and Orville-Thomas, W. J., *Molecular Interactions*, John Wiley, NY, 1982, vol. 3.
18. Szczepaniak, K. and Tramer, A., *J. Phys. Chem.*, 1967, **71**, 3035.
19. Zeegers-Huyskens, Th., *J. Mol. Struct.*, 1975, **26**, 329.
20. Shanmugasundaram, V. and Meyyappan, M., *Z. Phys. Chem. (Leipzig)*, 1977, **258**, 673.
21. Singh, A., Misra, R., Shukla, J. P. and Saxena, M. C., *Indian J. Pure Appl. Phys.*, 1983, **21**, 228.
22. Sabesan, R., Varadarajan, R. and Sargurumoorty, M., *Indian J. Phys.*, 1981, **B55**, 353.
23. Boobyer, G. J. and Orville-Thomas, W. J., *Spectrochim. Acta*, 1966, **22**, 147.

Received 8 July 2003; revised accepted 19 November 2003

Bioinformatic analysis of the SARS virus X1 protein shows it to be a calcium-binding protein

Abhishek Narain Singh[†], Dinesh Gupta[#] and Shahid Jameel[‡,*]

[†]Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, New Delhi 110 017, India
[#]Bioinformatics Facility and [‡]Virology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India

The X1 open reading frame of the SARS virus encodes a potentially unique protein not found in other coronaviruses. Here we present results of bioinformatic analyses that predict the X1 protein to bind calcium. The regulation of calcium homeostasis and calcium-mediated signalling pathways in infected cells may affect cell fate and contribute to the unique pathogenesis of this virus.

THE SARS coronavirus (SARS-CoV) was identified as the aetiological agent for severe acute respiratory syndrome (SARS), a form of atypical pneumonia that recently infected over 8000 persons and caused 916 deaths (http://www.who.int/csr/sars/en/). While the infection and disease was largely concentrated in China, Hong Kong, other parts of Southeast Asia and Canada, many other

*For correspondence. (e-mail: shahid@icgeb.res.in)

countries implemented quarantine measures and travel restrictions to contain its spread.

Coronaviruses are members of a family of enveloped viruses that replicate in the cytoplasm of animal host cells[1]. They are distinguished by the presence of a single-stranded plus sense RNA genome of 29–31 kb that has a 5′ cap and a 3′ polyA tract. The genome of SARS-CoV was rapidly sequenced from multiple geographic isolates[2–4] (www.ncbi.nlm.nih.gov), setting the stage for further research into molecular aspects of viral biology and pathogenesis. Besides genes found in all coronaviruses, the SARS-CoV was found to contain some unique genes. We applied various bioinformatics tools to analyse those unique genes with the potential to code for polypeptides larger than 50 amino acids. Here we describe our results with the *X1* gene[3], also called *ORF3* (ref. 2).

The *X1* gene spans nucleotides 25,268 to 26,092 (Tor2 isolate) and is predicted to encode a protein of 274 amino acids containing three transmembrane domains and a 153 amino acid C-terminal cytoplasmic domain[3]. BLASTP analysis (http://www.ncbi.nlm.nih.gov:80/BLAST/) with the entire X1 sequence showed weak alignments to a calcium pump from *Plasmodium falciparum* (NCBI: NP_703265, 1923410A and CAD58779.1) and glutamine synthetase from *Leptospira interrogans* (NCBI: NP_711494.1). The similarity was restricted to amino acid residues 209–264 in the cytoplasmic domain of X1. Since structures of the *P. falciparum* calcium pump have not been determined, we carried out pairwise alignment of the X1 sequence to calcium pumps from *Xenopus laevis* (1CFF_A) and *Oryctolagus cuniculus* (1EUL_A), and to the calcium-binding protein calmodulin (1NWD_A) whose three-dimensional structures were available in the Protein Databank (PDB; www.rcsb.org). Using ClustalW (www.ebi.ac.uk/clustalw/), both pumps showed good alignment to the 214–239 region of the X1 protein. However, the similarity with calmodulin was spread over the entire cytoplasmic domain of X1 (Figure 1). This encouraged us to further explore possible calcium-binding motifs in X1. The glutamine synthetase structure (2LGS) was similarly analysed, but no similarity to the X1 cytoplasmic domain was found within the Mn^{2+}-binding pocket of glutamine synthetase.

The Transmembrane Hidden Markov Model (TMHMM) software had predicted the cytoplasmic domain of X1 to span amino acids 122–274. For possible calcium-binding motifs, we focused on this region. The X1 sequence 118-243 showed optimal alignment to Calmodulin (1NWD_A) (Figure 1). This region was then used to generate a three-dimensional model for X1 through alignment interface on the Swiss-Model server (www.expasy.org/swissmod/) using the Calmodulin (1NWD_A) structure as a template. This is shown in Figure 2. The energy of this model is –1848.435 KJ/mole. According to PROCHECK analysis, the Ramachandran plot and other parameters were within standard acceptable limits for a model based on templates where very little similarity exists between the template and the model. While there are four calcium ions bound in calmodulin, based on sequence alignment, this model predicts the X1 cytoplasmic domain to bind two calcium ions. The calcium-binding pockets in the X1 cytoplasmic domain show the 'EF hand' motif commonly found in many calcium-binding proteins. Though three regions of X1 were predicted by CLUSTALW to bind Ca^{2+}, the model picked only two of these as calcium-binding folds. This was based on percentage similarity and the coordination chemistry of calcium and provides further confidence for our model.

In a complementary approach, we carried out structural alignment of the X1 protein through Metaserver analysis (http://bioinfo.pl/meta)[5] ID = 16866. The Metaserver analysis works at two levels. Initially it takes the query sequence (X1 in our case) and predicts a secondary structure based on various algorithms to derive a consensus. For X1, the Metaserver analysis used the sam-t99-2d, sam-t02-dssp, sam-t02-stride and profsec algorithms. At the next level, the predicted X1 secondary structure was matched to the existing database of protein secondary structures to give match scores. The Metaserver results obtained were individually analysed for structures that corresponded to metal-binding proteins. While the metal-binding regions in many proteins showed structural similarity to the cytoplasmic domain of X1, the best match

Figure 1. Alignment of X1 with calcium-binding proteins. The alignment was based on ClustalW results for calmodulin (1NWD_A), *X. laevis* calcium pump (1CFF_A) and *O. cuniculus* calcium pump (1EUL_A). The regions of similarity in X1 and calmodulin are indicated between the two sequences. Structural alignments of X1 from Metaserver results are also shown for endo/exocellulase catalytic domain (1TF4_A), human 4-sulfatase (1FSU) and the receptor-binding protein P2 of bacteriophage Prdl (1N7V_A). For clarity, only the relevant residues in these proteins have been shown.

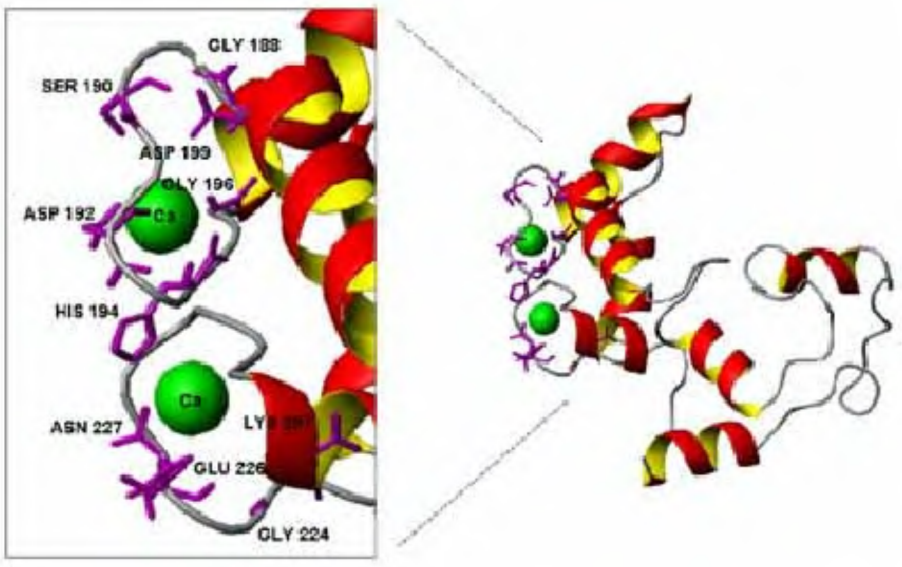

Figure 2. A 3D model of the cytoplasmic domain of the SARS virus X1 protein. This model is based on ClustalW alignment of the 118–243 region of X1 and was generated using the Swiss-Model server (www.expasy.org/swissmod/SWISS_MODEL.html) through the Alignment Interface. The two calcium ions and the surrounding residues in X1 are indicated.

was observed with calcium-binding proteins, strengthening our BLASTP analysis.

It has been suggested that the X1 protein might encode a domain with ATP-binding properties (PD037277)[2]. In a PROSITE scan (www.expasy.ch/prosite), though ATP-binding domains were found in the calcium pump (PS00154) and glutamine synthetase (PS00181), none was detected in the X1 sequence. An InterPro scan (www.ebi.ac.uk/interpro/scan.html) also did not reveal any ATP-binding domain. However, it showed a domain of undefined function called DUF109, to be present in X1 (residues 160–180) as well as in 13 other proteins; all of these proteins are hypothetical with no assigned function. Our analysis suggests that X1 is likely to be a calcium-binding protein with no ATP-binding property or ATPase function.

Calcium is a versatile signal that regulates many cellular functions[6], including cell survival and apoptosis[7]. Viruses can subvert cellular pathways to promote their survival and replication. Indeed, calcium homeostasis is known to play a role in the replication and pathogenesis of many viruses[8]. The intracellular levels of calcium are controlled through multiple sensory proteins at the level of plasma membrane and the endoplasmic reticulum, the latter being the major intracellular calcium store[9]. As it is predicted to be an integral membrane protein, X1 is likely to be distributed at the plasma membrane and the endoplasmic reticulum membrane, with its C-terminal domain in the cytoplasm. Thus, the protein would be ideally placed to regulate calcium homeostasis in an infected cell. Multiple calcium-binding cellular proteins are also involved in signalling pathways that regulate cell fate. We propose this role for the X1 protein in SARS pathogenesis.

1. Lai, M. M. C. and Holmes, K. V., Coronaviridae: The Viruses and their Replication. In *Field's Virology* (eds Knipe, D. M. and Howley, P. M.), Lippincott, Williams and Wilkins, Philadelphia, 2001, vol. 1, pp. 1163–1185.
2. Marra, M. A. *et al.*, The genome sequence of SARS-associated coronavirus. *Science*, 2003, **300**, 1399–1404.
3. Rota, P. A. *et al.*, Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003, **300**, 1394–1399.
4. Ruan, Y. J. *et al.*, Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet*, 2003, **361**, 1779–1785.
5. Ginalski, K. *et al.*, 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, 2003, **19**, 1015–1018.
6. Berridge, M. J. *et al.*, The versatility and universality of calcium signaling. *Nat. Rev. Mol. Cell. Biol.*, 2000, **1**, 11–21.
7. Orrenius, S. *et al.*, Regulation of cell death: the calcium-apoptosis link. *Nat. Rev. Mol. Cell. Biol.*, 2003, **4**, 552–565.
8. Ruiz, M. C. *et al.*, Role of Ca^{2+} in the replication and pathogenesis of rotavirus and other viral infections. *Cell Calcium*, 2000, **28**, 137–149.
9. Berridge, M. J. *et al.*, Calcium signaling: dynamics, homeostasis and remodeling. *Nat. Rev. Mol. Cell. Biol.*, 2003, **4**, 517–529.

Received 27 October 2003; revised accepted 11 February 2004