Acknowledgements

This work was supported by a grant from Natural Science and Engineering Research Council (NSERC) of Canada. R. Cedergren is fellow of the Evolutionary Biology programme of the Canadian Institute for Advanced Research. V. Bourdeau holds a predoctoral fellowship from NSERC.

References

1 Uhlenbeck, O. C. (1995) *RNA* 1, 4–6
2 Laferrière, A., Gautheret, D. and Cedergren, R. (1994) *CABIOS* 10, 211–212
3 Steinberg, S. and Cedergren, R. (1995) *RNA* 1, 886–891
4 Weeks, K. M. and Crothers, D. M. (1991) *Cell* 66, 577–588
5 Moore, P. S. *et al.* (1996) *J. Virol.* 70, 549–558
6 Gompels, U. A., Craxton, M. A. and Honess, R. W. (1988) *J. Virol.* 69, 2819–2829
7 Chang, Y. *et al.* (1994) *Science* 266, 1865–1869
8 Vogel, J. *et al.* (1988) *Nature* 335, 606–611
9 Karn, J. and Graeble, M. A. (1992) *Trends Genet.* 8, 365–368
10 McMillan, N. A. *et al.* (1995) *Virology* 213, 413–424
11 Cribier, B. *et al.* (1995) *AIDS* 9, 1131–1136
12 Ellington, A. D., Leclerc, F. and Cedergren, R. (1966) *Nat. Struct. Biol.* 3, 981–984

GERARDO FERBEYRE, VÉRONIQUE BOURDEAU AND ROBERT CEDERGREN

Département de Biochimie, Université de Montréal, Montréal, Québec, Canada H3C 3J7.
Email: ceder@bch.umontreal.ca

Enzyme HIT

We have discovered a striking resemblance of three-dimensional folds, despite low sequence identity, between the 'histidine triad', alias HIT family of proteins[1] and galactose-1-phosphate uridylyltransferase, GalT. Analysis of sequence and structure conservation leads to the evolutionary unification of the two families, and to detailed functional predictions for HIT proteins.

Members of the HIT protein family are ubiquitous in nature and believed to have an important cellular role that is reflected in their strong sequence conservation (e.g. 72% identity between human and prokaryotic *Haemophilus influenzae* sequences). However, the precise function *in vivo* of HIT proteins is unknown, except for two remotely related branches: diadenosine 5′,5‴-P^1,P^4-tetraphosphate (Ap_4A) asymmetrical hydrolase and Ap_4A phosphorylase[2].

The Dali/FSSP classification of three-dimensional structures[3] reveals extensive similarities indicative of a common evolutionary origin between a member of the HIT family, protein kinase C-interacting protein (PKCI-1; Ref. 4) and GalT[5]. Fortunately, the function of GalT is known: the enzyme transfers UDP between galactose- and glucose-1-phosphate, thus balancing the cellular pools of biochemically activated sugars. The structural surprise is that the three-dimensional chain trace of GalT (334 residues) corresponds to two sequentially fused PKCI-1 protomers (113 residues each) conserving the quaternary structure of the PKCI-1 dimer and leading to superimposition of 198 pairs of equivalent Cα atoms with 3.1 Å positional root-mean-square deviation. Accordingly, there is an intramolecular dyad axis of approximate symmetry in GalT that corresponds to the intermolecular dyad axis of the PKCI-1 dimer. The PKCI-1 protomer and the two subdomains of GalT appear to be three incarnations of an ancient functional core (Figs 1, 2a–c, key residue positions labelled *a–h*).

The striking structural similarity can be exploited to generate a hypothesis about common function. It is plausible that the GalT and HIT protein families represent a structurally and mechanistically conserved enzyme superfamily with specialized substrate specificities. In GalT, the histidine at site *g* of the zinc-binding domain acts as the catalytic nucleophile that transiently binds UMP during the hexose-phosphate transfer reaction. The histidine at site *g* is the only residue that is invariantly conserved through all of the Ap_4A phosphorylase, Ap_4A hydrolase, HIT and GalT sequences. We predict that cleavage of nucleotidyl substrates at the β-phosphate by a mechanism involving a covalent phospho-histidine intermediate at site *g* is the common biochemical function of all four enzyme subgroups, including HIT proteins (Fig. 2b). The conformation of loops lining the binding pocket is remarkably well-conserved, which is evidence of strong functional (rather than structural) constraints, and a number of substrate-binding residues are conserved. These observations support the hypothesis that the HIT family and GalT bind chemically similar substrates. For example, we expect Ap_4A hydrolase to bind part of Ap_4A at a location equivalent to that of UDP in GalT.

The histidine pattern of HIT proteins was originally implicated in zinc binding[1,6]. In light of the known three-dimensional structures and the scarcity of invariantly conserved residues in the multiple alignment of currently known member sequences of the superfamily, the role of metal binding becomes controversial. Most importantly, the zinc at site *a-b-d-f* is not directly involved in the

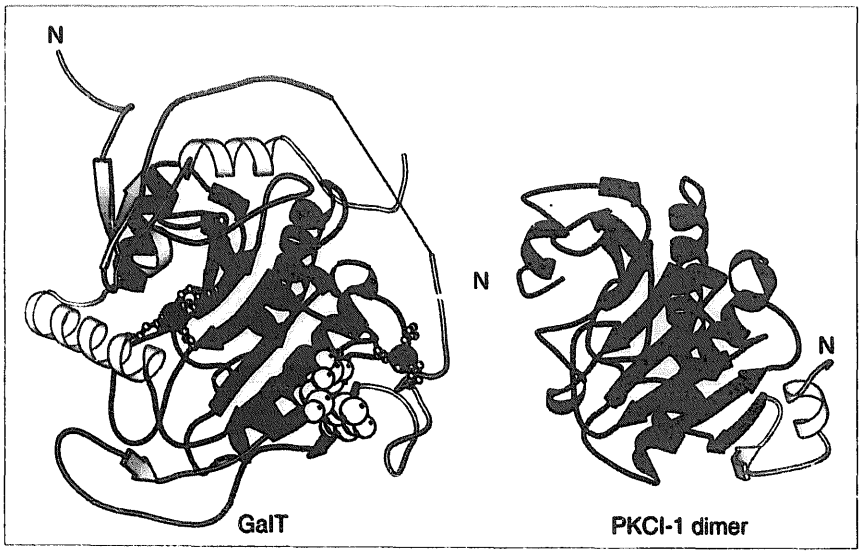

Figure 1
Structure comparison between galactose-1-phosphate uridylyltransferase (GalT) and protein kinase C-interacting protein (PKCI-1), a member of the 'histidine triad' or HIT family of proteins. Intramolecular duplication revealed by structural alignment. The same structural motif repeats twice in the PKCI-1 homodimer (green and red-white protomers) and twice in the structure of GalT (green: iron-binding subdomains; red: zinc-binding subdomain). The connection between the duplicated subdomains in GalT is made by the white helix at the bottom left. On top of the common core, GalT has acquired new structural elements involved in dimerization (yellow loops). The UMP substrate of GalT is bright yellow. Sidechains are shown for metal ligands of GalT. Plotted with Molscript[7].

 0968–0004/97/$17.00 PII: S0968-0004(97)01021-9

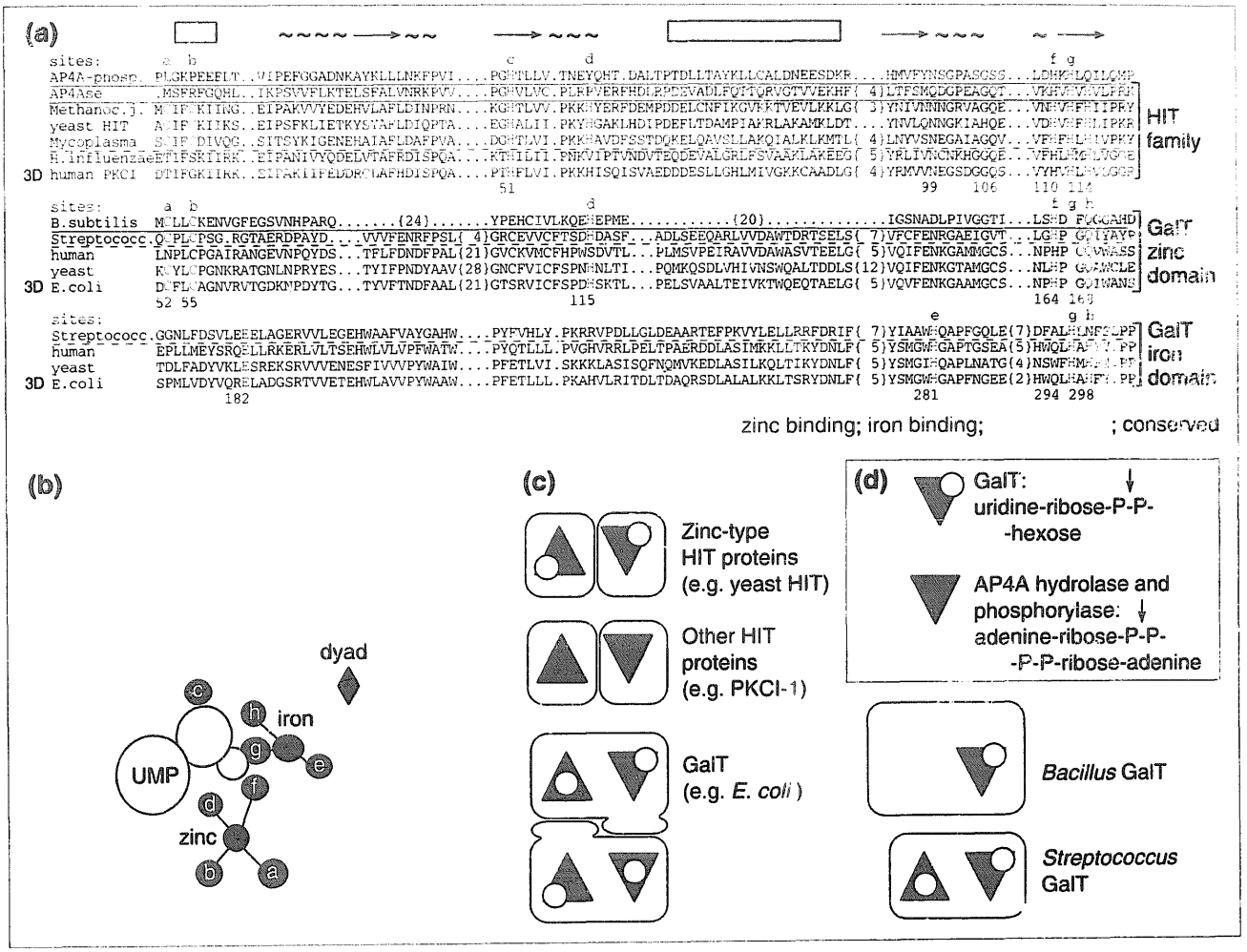

Figure 2

Structure comparison between galactose-1-phosphate uridylyltransferase (GalT) and protein kinase C-interacting protein (PKCI-1), a member of the 'histidine triad' or HIT family of proteins. Correlated mutations of functional residues. **(a)** A striking set of correlated mutations involving sites *a-b-d-f* suggest that a subset of HIT proteins have a similar zinc-binding site to the amino-terminal domain of GalT. The alignment shows representative sequences aligned within families by Clustalw[8] and between families based on the structural alignment by Dali[9]. Horizontal lines divide the families into subfamilies on the basis of sequence similarities. GalT from *Bacillus subtilis* belongs to a manually aligned outgroup that appears to contain a single copy of the zinc-binding domain embedded in a large protein of about 500 residues. Sites *f-g-h* are the histidine triad after which the HIT family of proteins was named[1]; the more-recently sequenced PkiA protein from *Dictyostelium* shows that site *f* is variable (data not shown). Structural conservation is indicated at the top as: helix, striped rectangles; strand, arrows; and loop, squiggles. Residue numbers refer to the crystal structures of human PKCI-14 and *Escherichia coli* GalT[5]. **(b)** Schematic location of functional residues identified from the crystal structure of GalT[5] shown in a superimposition of the iron- and zinc-binding domains. Structurally equivalent residues are labelled as in the multiple alignment. **(c)** Enzyme architectures. Triangles denote the functional core. Phylogenetic groupings are indicated by colour and known or predicted metal-binding sites by circles. The *Streptococcus* GalT is depicted as a monomer as it seems to lack several elements of the dimer interface despite 35% residue identity with the *E. coli* sequence. **(d)** Known catalytic activities: arrows indicate site of attack. Colouring refers to phylogenetic groupings in (c).

bond formation or cleavage step of GalT[6] and, indeed, this zinc site is lost from human GalT.

Surprisingly, a subset of HIT proteins (including the original HIT gene from yeast) nevertheless do contain the GalT-type zinc-binding site *a-b-d-f* (correlated mutation in Fig. 2a), leading to the prediction that this particular subset of HIT enzymes is active in the presence of bound zinc. However, the PKCI-1 protein does not share the *a-b-d-f* signature, but there are reports of zinc binding leading to conformational changes and increased susceptibility to proteolysis[4]. We, therefore, propose that the zinc-free form of PKCI-14, is the biologically active conformation of this enzyme.

In summary, comparative structure/sequence analysis leads to precise hypotheses concerning the function of HIT proteins. We shall eagerly await verification (or falsification) of these functional predictions by binding assays and site-directed mutagenesis.

References

1 Séraphin, B. (1992) *DNA Seq.* 3, 177–179
2 Huang, Y., Garrison, P. N. and Barnes, L. D. (1995) *Bicchem. J.* 312, 925–932
3 Holm, L. and Sander, C. (1996) *Nucleic Acids Res.* 24, 206–210
4 Lima, C. D., Klein, M. G., Weinstein, I. B. and Hendrickson, W. A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 5337–5362
5 Wedekind, J. E., Frey, P. A. and Rayment, I. (1995) *Biochemistry* 34, 11049–11061
6 Arabshahi, A., Ruzicka, F. J., Geeganage, S. and Frey, P. A. (1996) *Biochemistry* 35, 3426–3428
7 Kraulis, P. (1991) *J. Appl. Crystallogr.* 24, 946–950
8 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) *Nucleic Acids Res.* 22, 4673–4680
9 Holm, L. and Sander, C. (1993) *J. Mol. Biol.* 233, 123–138

LIISA HOLM AND CHRIS SANDER

EMBL–EBI, Wellcome Trust Genome Campus, Cambridge, UK CB10 1SD.