Protein Science (1994), *3*:1128–1130. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Structural similarity between binding sites in influenza sialidase and isocitrate dehydrogenase: Implications for an alternative approach to rational drug design

ANDREW R. POIRRETTE,[1] PETER J. ARTYMIUK,[2] HELEN M. GRINDLEY,[1] DAVID W. RICE,[2] AND PETER WILLETT[2]

[1] Department of Information Studies and [2] Department of Molecular Biology and Biotechnology, Krebs Institute, Sheffield University, Sheffield S10 2TN, United Kingdom

(RECEIVED December 23, 1993; ACCEPTED April 26, 1994)

Abstract: Using searching techniques based on algorithms derived from graph theory, we have established a similarity between a 3-dimensional cluster of side chains implicated in drug binding in influenza sialidase and side chains involved in isocitrate binding in *Escherichia coli* isocitrate dehydrogenase. The possible implications of the use of such comparative methods in drug design are discussed.

Keywords: binding sites; drug design; graph theory; influenza sialidase; isocitrate dehydrogenase; structure similarity

An important factor in the development of molecular biology has been the availability of computational tools for the detection of similarities in the 1-dimensional sequences of proteins and nucleic acids (Lesk, 1988). It is similarly of the greatest importance to be able to detect structural analogies within the rapidly growing database of 3-dimensional protein and nucleic acid structures in order to enhance understanding of structure/function relationships in biological macromolecules. In this communication we describe the use of a novel program, ASSAM (Artymiuk et al., 1994), which uses algorithms derived from graph theory, in order to establish a thought-provoking analogy between binding sites in the active sites of influenza sialidase and isocitrate dehydrogenase. This analysis suggests that such comparative studies may represent another valuable approach to the rational design of novel inhibitors.

Recently, von Itzstein et al. (1993) described the design of potential anti-influenza drugs, which operate by binding to a specific site on the influenza sialidase molecule, thereby inhibiting it. This was achieved using a combination of computer-assisted manual examination of the active-site structure with sialic acid and sialic acid analogues bound (Chong et al., 1992; Varghese et al., 1992), together with the use of the GRID program (Goodford, 1985) to explore probable interaction sites between probes with various functional group characteristics and the enzyme surface. A key feature of their analysis of the sialidase structure was the recognition of a negatively charged patch on the sialidase surface provided by 2 glutamate residues and not used in the binding of sialic acid. This patch was exploited in the rational design process by the synthesis of a sialic acid analogue bearing an additional positive charge from either an amino or a guanidinyl group appropriately placed to interact with this region of the sialidase surface. This work constituted a long-awaited example of rational computer-assisted design of a new drug based on the crystal structure of a target protein (Taylor, 1993) and highlighted the great importance of knowing the 3-dimensional structures of medically important macromolecules.

In earlier work (Mitchell et al., 1990; Grindley et al., 1993), we described computer programs that use algorithms derived from graph theory to detect similarities in protein folds at the level of helices and strands. We have now extended this work to compare arrangements of side chains in 3-dimensional space. A simplified representation, consisting of pseudo-atoms at the beginning and end of the functional part of each side chain, is used for each side chain in each protein in the database and for the search pattern. Within user-specified tolerances, the ASSAM program compares the search pattern with the database proteins using the subgraph isomorphism algorithm of Ullmann (1976). Suitable search patterns consist of small groups of side chains and searches are very rapid, typically requiring about 2 min CPU time on an R4000 Silicon Graphics Indigo.

Von Itzstein et al. (1993) identified a group of 4 side chains as being involved in drug binding in the influenza sialidase (2 arginines, 371 and 118, and 2 carboxylic acid groups, glutamates 119 and 227). Coordinates for these side chains were taken from the sialidase structure (Protein Data Bank [PDB] accession no. 1NN2; Varghese & Colman, 1991) and were used to generate a search pattern corresponding to this side-chain cluster. Using the ASSAM program, we have compared this 3-dimensional arrangement of side chains with the other protein structures in the PDB (October 1992 release; Bernstein et al., 1977). The search

Reprint requests to: Peter J. Artymiuk, Department of Molecular Biology and Biotechnology, Krebs Institute, Sheffield University, Sheffield S10 2TN, UK; e-mail: mb1pja@sunc.sheffield.ac.uk.

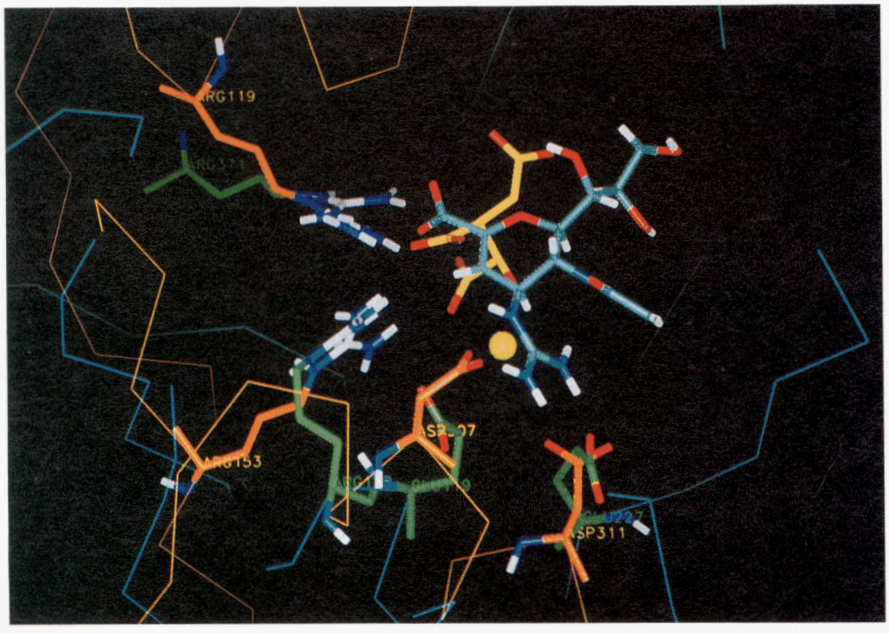

Fig. 1. Superposition of the sialidase and ICDH binding sites. Arginines 371 and 118 and glutamates 119 and 227 from sialidase are shown with green carbon atoms (from PDB coordinate set 1NN2; Varghese & Colman, 1991). The inhibitor 4-guanidino-Neu5Ac2en has been modeled in position after von Itzstein et al. (1993) and is shown with blue-green carbon atoms. Arginines 119 and 153 and aspartates 307 and 311 of ICDH are shown with orange carbon atoms, the isocitrate with yellow carbon atoms, and the Mg^{2+} ion as a yellow sphere (coordinates from PDB coordinate set 5ICD; Hurley et al., 1990). Nitrogen, oxygen, and hydrogen atoms are shown in dark blue, red, and white, respectively. This diagram was produced using the Sybyl program (Tripos Associates Inc., St. Louis, Missouri).

revealed that there is a very similar cluster of 4 side chains in the the binding pocket for the isocitrate/Mg^{2+} complex in the active site of *Escherichia coli* isocitrate dehydrogenase (ICDH; Hurley et al., 1990), an enzyme that catalyzes the $NADP^+$-linked oxidative decarboxylation of isocitrate to 2-oxoglutarate.

The 2 sets of side chains from sialidase and from ICDH are shown superposed in Figure 1 and Kinemage 1, where it can be seen that in ICDH the positions of arginines 119 and 153 correspond closely to arginines 371 and 118, respectively, in the sialidase, with excellent overlap of the guanidinium groups. In ICDH these 2 arginines interact with 2 carboxyl groups of the isocitrate, one of which, the C3 carboxyl, occupies an equivalent position to the carboxyl group of the inhibitor 4-guanidino-Neu5Ac2en in sialidase. The 2 carboxyl groups in the ICDH site are provided by 2 aspartates (aspartates 307 and 311), which are slightly displaced with respect to their counterparts in sialidase (glutamates 119 and 227, respectively). In sialidase these carboxyls coordinate the positive charge of the 4-guanidinyl group of the inhibitor, and in ICDH the equivalent carboxyls analogously coordinate the positive charge of the Mg^{2+} ion in the complex.

Clearly, this result does not mean that the sialidase would necessarily be expected to bind isocitrate. The binding site for sialic acid in sialidase comprises many more residues than the 2 arginines and 2 aspartates that were highlighted by von Itzstein et al. (1993) and which we used as our search pattern. In fact, these 4 residues constitute the only major area of resemblance between the binding sites in the 2 proteins; the folds of the polypeptide chain and the positions of other residues in the active site are quite different. Thus, the result reported here does not imply that an isocitrate/Mg^{2+} complex would necessarily bind to the sialidase. Nevertheless, this finding does suggest that compounds with an isocitrate-like framework and a similar pattern of charge distribution might form a suitable alternative starting point for the design of new families of anti-influenza drugs.

In more general terms, it is clear that structural comparisons of the kind reported here may be more widely applicable. At present, the database of known 3-dimensional protein structures is very small, consisting of only a few hundred distinct structures. As this database expands, it is increasingly likely that more similarities of the kind we have described above will be observed between otherwise disparate proteins. At the scientific level, such similarities will enhance our understanding of structure/function relationships in proteins while technologically they may prove valuable both in protein engineering and in the search for new lead compounds in drug design. Thus, it seems likely that structural comparisons will be a useful weapon to be used in conjunction with other modeling procedures.

Acknowledgments

This work was supported by the Science and Engineering Research Council, The Wellcome Trust, the Lister Institute, and Tripos Associates Inc. P.J.A. is a Royal Society University Research Fellow. The Krebs Institute is a designated SERC Biomolecular Sciences Centre.

References

Artymiuk PJ, Grindley HM, Poirrette AR, Rice DW, Ujah EC, Willett P. 1994. Identification of β-sheet motifs, of ψ-loops, and of patterns of amino acid residues in three-dimensional protein structures using a subgraph isomorphism algorithm. *J Chem Inf Comput Sci 34*:54–62.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi M, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Chong AKJ, Pegg MS, Taylor NR, von Itzstein M. 1992. Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur J Biochem 207*:335–343.

Goodford PJ. 1985. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J Med Chem 28*:849–857.

Grindley HM, Artymiuk PJ, Rice DW, Willett P. 1993. Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J Mol Biol 229*:707–721.

Hurley JH, Dean AM, Sohl JL, Koshland DE Jr, Stroud RM. 1990. Regu-

lation of an enzyme by phosphorylation at the active site. *Science 249*:1012–1016.
Lesk AM. 1988. *Computational molecular biology*. Oxford, UK: Oxford University Press.
Mitchell EM, Artymiuk PJ, Rice DW, Willett P. 1990. Use of techniques derived from graph theory to compare secondary structure motifs in proteins. *J Mol Biol 212*:151–166.
Taylor GL. 1993. A rational attack on influenza. *Nature 363*:401–402.
Ullmann JR. 1976. An algorithm for subgraph isomorphism. *J Assoc Computing Machinery 23*:31–42.
Varghese JN, Colman PM. 1991. 3-dimensional structure of the neuraminidase of influenza-A / Tokyo / 3 / 67 at 2.2 Å resolution. *J Mol Biol 221*:473–486.
Varghese JN, McKimm-Breschkin J, Caldwell JB, Korrt AA, Colman PM. 1992. The structure of the complex between the influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins Struct Funct Genet 14*:327–332.
von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, Phan TV, Smythe ML, White HF, Oliver SW, Colman PM, Varghese JN, Ryan DM, Woods JM, Bethell RC, Hotham VJ, Cameron JM, Penn CR. 1993. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature 363*:418–423.