FOR THE RECORD

MIF proteins are not glutathione transferase homologs

WILLIAM R. PEARSON

Department of Biochemistry, University of Virginia, Charlottesville, Virginia 22908

(RECEIVED January 14, 1994; ACCEPTED January 19, 1994)

Abstract: Although macrophage migration inhibitory factor (MIF) proteins conjugate glutathione, sequence analysis does not support their homology to other glutathione transferases. Glutathione transferases are not detected with MIF proteins in searches of protein sequence databases, and MIF proteins do not share significant sequence similarity with glutathione transferases. Homology cannot be demonstrated by multiple sequence alignment or evolutionary tree construction; such methods assume that the proteins being analyzed are homologous.

Keywords: glutathione transferase; homology; inhibitory factor; macrophage migration; sequence similarity

Macrophage migration inhibitory factor proteins (MIF) have been shown to possess glutathione transferase activity (Blocki et al., 1992) and share 28% N-terminal sequence identity (7/25) with class- μ glutathione transferases. Based on these observations and on evolutionary tree analysis using the PIMA (pattern-induced multiple alignment) program (Smith & Smith, 1992), Blocki et al. (1993) have argued that MIF proteins are homologous to the soluble glutathione transferase family and are most closely related to the theta class (Pemble & Taylor, 1992). However, multiple alignment and evolutionary tree programs cannot be used to demonstrate sequence homology, because these programs assume that the sequences being considered are homologous. To demonstrate from sequence data the likelihood of homology, one must use programs that compare sequences to protein databases or evaluate the statistical significance of similarity scores. Based on similarity scores and sequence alignments, there is no evidence that MIF and glutathione transferases share a common ancestor.

Three lines of evidence argue against common ancestry for the soluble glutathione transferases and the MIF proteins. (1) A search of the protein sequence databases with the MIF sequence does not reveal similarities to members of the glutathione transferase family. (2) The level of similarity shared between MIF and other glutathione transferases is no higher than would

be expected from a random sequence with similar amino acid composition. (3) Except for proteins where exon shuffling has occurred, homologous proteins have similar length and are usually similar across their entire length. MIF proteins are about 115 residues long, whereas glutathione transferases are about 100 residues longer. Similarities between MIF and glutathione transferases extend from 20 to 100 residues, but typically align portions of MIF with different regions in different glutathione transferase classes.

Table 1A shows similarity scores, statistical significance, percent identities, length of alignment, and alignment phasing for alignments between a MIF protein, mouse delayed early response gene 6 (DER6), and representative glutathione transferases. These similarity scores are very low: the glutathione transferase with the highest similarity to DER6 has a score of 58. A score this high or better would be expected to be seen more than 800 times if a random sequence of the same length and amino acid composition were used for the search. In contrast, when the rat class- θ sequence is compared to a representative set of glutathione transferase sequences (Table 1B), all the soluble mammalian enzymes have scores with E-values less than 50, while the unrelated microsomal glutathione transferases and MIF proteins have scores that are only slightly above average. Note also that even when very distantly related glutathione transferases are compared, the N-terminal sequences of both proteins are aligned (Table 1B, "Off" column). In contrast, the MIF protein matches very different parts of different glutathione transferases, as would be expected from a random alignment. Thus, while the 28% N-terminal identity (7/25) may appear striking at first glance, such levels of similarity are quite common between unrelated proteins. Yeast *N*-acetyl gamma-glutamyl-phosphate reductase shares 2 regions of similarity: one with 30% identity (12/40) and 50% similarity (16/40) and a second with 52% identity (9/16) and 64% similarity (11/17).

Multiple alignment and evolutionary tree construction programs cannot be used to demonstrate sequence homology; these programs assume a priori that the sequences being examined are related. Thus, one can just as easily build a tree that includes human alpha-globin and the glutathione transferases as with a MIF protein. Such trees can be very misleading, because frequently they will cluster the unrelated sequence with the sequence family member that is most divergent from the rest of the fam-

Reprint requests to: William R. Pearson, Department of Biochemistry, Box 440, Jordan Hall, University of Virginia, Charlottesville, Virginia 22908; e-mail: wrp@virginia.edu.

Table 1. Similarity scores, statistical significance, percent identities, length of alignment, and alignment phasing for alignments between MIF proteins and representative glutathione transferases

Locus	Description	Len	Score ^a	E ^b	% ^c	ALen ^c	Off ^d
A. Comparison with mouse DER6							
A44499	Mouse DER6	115	561	10 ⁻³⁶	100.0	115	1
A33838	Human MIF	115	539	10 ⁻³⁵	88.7	115	1
S27117	Rat 13K MIF (N-term.)	26	120	10 ⁻²	100.0	26	1
A29036	Rat class-mu Yb3 GST	218	58	871	16.2	105	60
S03015	Human class-pi GST	210	56	1,233	15.4	91	52
A26484	<i>S. japonicum</i> 28K	218	55	1,466	18.9	95	106
JS0618	Rat class-theta GST	244	48	4,664	22.2	63	121
S14636	Chicken CL2 GST	220	47	5,443	16.9	83	60
S16293	House fly GST	208	38	17,254	20.0	20	-10
S03615	<i>C. elegans</i> GST	208	37	18,929	25.5	55	-4
XUFF11	Fruit fly GST	209	37	18,929	16.7	18	-10
A26753	Rat class-alpha GST	221	34	23,732	20.0	20	-92
XUZM32	Maize GSTIII	222	34	23,732	24.2	33	-13
DCMA_METSP	Dichloromethane dehalo.	288	32	26,467	16.7	18	-33
A23883	Rat microsomal GST	155	29	29,554	18.8	16	-85
B. Comparison with rat class-theta Yrs							
JS00618	Rat class-theta GST Yrs	244	1,232	10 ⁻⁶⁶	100.0	244	-1
S27161	Rat class-theta GST 5	239	688	10 ⁻³⁷	54.0	215	-1
DCMA_METSP	Dichloromethane dehalo.	288	296	10 ⁻¹⁰	29.4	204	12
S14345	Rat class-theta (frag.)	114	280	10 ⁻⁹	59.6	99	1
JQ1378	Fruit fly GST	212	236	10 ⁻⁶	26.9	193	-2
XUZM32	Maize GST	222	227	10 ⁻⁵	27.2	202	2
S16293	House fly GST	208	223	10 ⁻⁵	25.7	191	-2
S14346	Rat class-theta (frag.)	109	206	10 ⁻⁴	59.5	79	1
XUFF11	Fruit fly GST	209	190	10 ⁻³	29.7	165	-1
A29036	Rat class-mu GST	218	129	1	20.9	206	1
A26653	Rat class-alpha GST	222	128	1	20.3	207	1
S14636	Chicken CL2 GST	220	104	24	23.0	152	6
S09316	Human class-pi GST	209	102	32	19.8	172	-3
A26484	<i>S. japonicum</i> GST	219	89	162	21.4	192	-5
S03615	<i>C. elegans</i> GST	208	85	267	19.1	230	1
A23883	Rat microsomal GST	155	51	12,730	17.2	122	-99
A33838	Human MIF	115	50	13,812	22.2	63	-120
A44499	Mouse DER6	115	48	16,078	22.2	63	-120
S27117	Rat MIF (N-term.)	26	23	32,795	36.4	22	-154

^a Optimal local similarity score using PAM250 matrix, -12 for an initial gap and -2 for each additional gap.

^b Expectation value: the number of sequences having this score or higher when searching a database of 33,329 entries (SWISS-PROT 27) expected by chance. An extreme value distribution was fitted to the distribution of similarity scores to estimate values of lambda and *K* (Karlin & Altschul, 1990), which were then used to estimate the probability of obtaining the score by chance. The estimation routine was provided by P. Green.

^c Percent identity over the aligned local region of length ALen.

^d The offset of the alignment of the first sequence into the second sequence.

ily. Because there is no significant similarity between the unrelated sequence and the family, the branch to that sequence will be placed randomly.

The only evidence that supports homology between MIF proteins and the soluble glutathione transferases is their copurification with the glutathione transferases on *S*-hexyl-glutathione affinity columns and activity on 1,2-epoxy-3-(*p*-nitrophenoxy) propane (7 μ mol min⁻¹/mg; Blocki et al., 1992). However, such activity is not sufficient to infer common ancestry; the ability to conjugate glutathione has arisen several times in evolutionary history. The microsomal glutathione transferases (155 amino acids) are other mammalian enzymes that conjugate glutathione but do not share common ancestry with the soluble glutathione transferases (DeJong et al., 1988).

While sequence similarity can be used very effectively to demonstrate common ancestry, it is much more difficult to prove lack of homology. However, there is no significant relationship between the sequences of MIF proteins and glutathione transferases. Thus, in the absence of structural information for the MIF proteins, there are no data to support the common ancestry of MIF and glutathione transferases.

Acknowledgments

W.R.P. is supported by research grants from the American Cancer Society and the National Library of Medicine. Code for estimating probabilities of optimal similarity scores was provided by P. Green, Washington University, St. Louis, Missouri.

References

- Blocki FA, Ellis LBM, Wackett LP. 1993. MIF proteins are theta class glutathione S-transferase homologs. *Protein Sci* 2:2095-2102.
- Blocki FA, Schlievert PM, Wackett LP. 1992. Rat liver protein linking chemical and immunological detoxification systems. *Nature* 360:269-270.
- DeJong JL, Morgenstern R, Jornvall H, DePierre JW, Tu CP. 1988. Gene expression of rat and human microsomal glutathione S-transferases. *J Biol Chem* 263:8430-8436.
- Karlin S, Altschul SF. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci USA* 87:2264-2268.
- Pemble SE, Taylor JB. 1992. An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences. *Biochem J* 287:957-963.
- Smith RF, Smith TF. 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modelling. *Protein Eng* 5:35-41.