

## Insilco analysis of functionally important residues in folate receptors

Kalidoss Ramamoorthy<sup>1</sup>, Sirisha Potala<sup>1</sup> and Rama Shanker Verma<sup>1,\*</sup>

<sup>1</sup>Department of Biotechnology, Indian Institute of Technology Madras, Chennai-600036, TN, India;  
Rama Shanker Verma\* - E-mail: vermars@iitm.ac.in; Phone: 091 44 2257 4109; Fax: 091 44 2257 4102; \*Corresponding author

received November 15, 2007; revised November 23, 2007; accepted December 11, 2007; published online December 28, 2007

### Abstract:

Lack of crystal structure data of folate binding proteins has left so many questions unanswered (for example, important residues in active site, binding domain, important amino acid residues involved in interactions between ligand and receptor). With sequence alignment and PROSITE motif identification, we attempted to answer evolutionarily significant residues that are of functional importance for ligand binding and that form catalytic sites. We have analyzed 46 different FRs and FBP sequences of various organisms obtained from Genbank. Multiple sequence alignment identified 44 highly conserved identical amino acid residues with 10 cysteine residues and 12 motifs including ECSPNLGPW (which might help in the structural stability of FR).

**Keywords:** folate receptors (FR); folate binding proteins (FBP); multiple sequence alignment; consensus sequence; conserved motifs; evolutionary trace (ET)

### Background:

Folate is the major source of one carbon moiety during DNA biosynthesis in various organisms. Internalization of the folic acid into cell is mediated by folate receptors (FRs) or folate binding proteins (FBPs) or by reduced folate carriers (RFCs). [1, 2] The FRs are clustered on the cell surface and associated with uncoated membrane invaginations known as caveolae. [3, 4] Folate binds to the externally oriented receptor and is followed by internalization. After binding and internalization, at low pH folate receptor disassociates from folate and is transported back to the cell surface through potocytosis. [2, 5, 6] Based on the tissue expression and affinity of binding towards folic acid and its various analogues into three major types, FR1 (FR- $\alpha$ ), FR2 (FR- $\beta$ ), FR3 (FR- $\gamma$ ). [7] FR1 (FR- $\alpha$ ) and FR2 (FR- $\beta$ ) are membrane bound GPI anchored and expressed in adult epithelial cells and placental (fetal) tissues respectively [8, 9, 10], FR3 (FR- $\gamma$ ) is a secretory protein as it lacks signal for GPI attachment [11] and is developmentally highly regulated with a restricted spatial (tissue specific) or temporal (time specific) expression pattern. [12] The presence of fourth rare type is also reported as FR4 (FR- $\delta$ ). [10] FR1 and FR2 differ not only on tissue expression but also in binding affinities and stereo specificities to various folate analogues. [10] It is been reported that, Folate receptors and Riboflavin binding proteins share more than ~30% sequence similarity. [13] In both cases, sequences exhibited 16 conserved cysteine residues that form eight bridges important for ligand binding. The N-terminal and centre portion of the sequences show very high similarity rather than C-terminal. [14] Here, we attempted to explore the important residues that may play an important

role in the binding of folic acid and various analogues. This may help us to find mechanism of binding and interactions of the folate receptor subtypes with folic acid and its analogues under divergent physiologic condition. This study will provide information about the differential efficacy of the compounds such as, Methotrexate (MTX) in relation to developing new strategy in relation to cancer therapy.

Multiple sequence alignments are often used to find out conserved sequence regions in a group of sequences. Here, we have used ClustalW multiple sequence alignment tool which produces biologically meaningful alignments of divergent sequences. Through multiple sequence alignment we can also find out motifs which are short conserved sequence among set of sequences. The detection of conserved residues will be useful in identifying the functionally important residue even in the absence of structural information.

### Methodology:

#### Dataset

We retrieved 140 folate related sequences by keyword search at GenBank [15]. We manually curated the dataset to remove redundant sequences. Thus, a refined dataset of 46 non-redundant sequences of different folate receptors and folate binding proteins were created. To study evolutionarily trend among conserved functional residues in these sequences (about 250 residues in length), we grouped them based on sequence source (12 different organisms). (See: supplementary material).

### Multiple sequence alignment

The 46 sequences are subjected to multiple sequence alignment as a first step to assess sequence conservation of single residue or motifs (residue stretch). Multiple sequence alignment was done using CLUSTAL-W. [16] Different parameters were tested and manual editing was performed wherever required to get significant alignment.

### Motif Identification

The conserved motifs identified by multiple sequence alignment are submitted to PROSITE web server to scan against existing signatures and identify motifs unique to folate receptor [17].

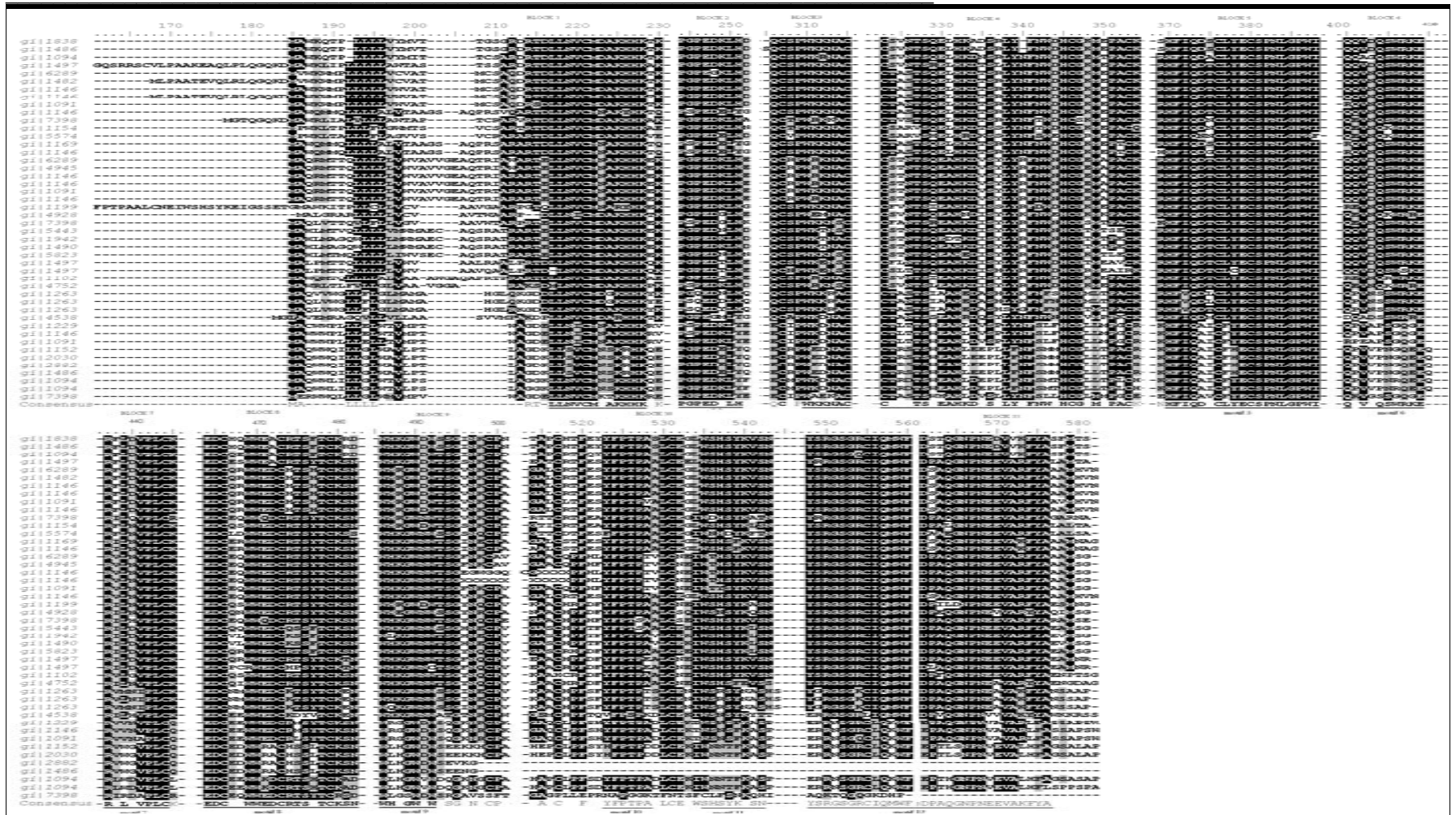
### Discussion:

#### Conserved and variable amino acids

Multiple alignments of all FBP sequences are presented in Figure 2. A list of conserved amino acids is presented in figure 1 with the percentage of conservation and the major substitutions present at the position. Figure 1 shows that L<sup>217</sup>, C<sup>221</sup>, M<sup>222</sup>, H<sup>227</sup>, K<sup>228</sup>, P<sup>244</sup>, E<sup>247</sup>, L<sup>250</sup>, C<sup>307</sup>, P<sup>309</sup>, W<sup>310</sup>, C<sup>315</sup>, C<sup>323</sup>, S<sup>328</sup>, F<sup>341</sup>, H<sup>345</sup>, C<sup>346</sup>, H<sup>370</sup>, F<sup>371</sup>, Q<sup>373</sup>, C<sup>376</sup>, E<sup>379</sup>, C<sup>380</sup>, S<sup>381</sup>, P<sup>382</sup>, N<sup>383</sup>, L<sup>384</sup>, G<sup>385</sup>, P<sup>386</sup>, W<sup>387</sup>, E<sup>409</sup>, R<sup>437</sup>, P<sup>442</sup>, L<sup>443</sup>, C<sup>444</sup>, E<sup>464</sup>, D<sup>465</sup>, C<sup>466</sup>, W<sup>470</sup>, C<sup>473</sup>, T<sup>478</sup>, C<sup>479</sup>, W<sup>486</sup> and W<sup>490</sup> residues are absolutely conserved in all the sequence of folate receptors. Although 44 residues showed 100% conservation, an additional 6 residues (K<sup>225</sup>, D<sup>248</sup>, T<sup>327</sup>, G<sup>347</sup>, I<sup>388</sup>, K<sup>480</sup>, and G<sup>489</sup>) showed more than 95% conservation (Figure 1).

Alignment position (a)	Residues (b)	Percentage (c)	Alignment position (a)	Residues (b)	Percentage (c)
218	L	100	372	I	94
				V	6
219	N	91	373	Q	100
	S	9			
220	V	91	376	C	100
	I	9			
221	C	100	377	L	79
222	N	100	379	F	21
	N	40		E	100
223	K	2	380	C	100
	D	58			
225	K	96	381	S	100
	R	4			
227	H	100	382	P	100
228	K	100	383	N	100
244	P	100	384	L	100
	G	81			
245	S	19	385	G	100
247	E	100	386	P	100
248	D	96	387	W	100
	G	4			
250	L	100	388	I	98
				F	2
251	H	70	409	E	100
	Y	30			
310	W	100	437	R	100
				V	35
313	N	90	438	I	33
	R	6		F	25
	K	4		L	4
314	A	85	442	P	100
	S	15			
315	C	100	443	L	100
323	C	100	444	C	100
327	T	96	465	D	100
	V	4			
328	S	100	466	C	100
336	S	87	470	W	100
	P	13		E	85
339	Y	87	471	R	9
	F	13		K	4
341	F	100	473	Q	2
				C	100
345	H	100	474	R	62
				K	24
346	C	100	478	H	10
	G	98		Q	4
347	S	2	479	T	100
348	M	77	480	C	100
	L	23		K	98
	K	77		R	2
353	R	15	482	N	85
	H	4		D	13
	E	2		S	2
	Q	2			
370	H	100	489	G	98
371	F	100	490	S	2
				W	100

**Figure 1:** Conserved amino acids in members of FBP sequences (a) Amino acid position by alignment (b) Consensus residues given at positions are in bold. Residues over 85% are shaded (c) Percentage of 46 aligned sequences in which the listed residue appears. Percentages are rounded to the nearest number. Residues over 85% are shaded



**Figure 2:** Shown in next four pages. Multiple sequence alignment of 46 members of FBPs. Conserved residues are highlighted in block (an arbitrary cut-off of 50% was used for conservation). Dashes represent gaps, introduced for the best alignment. A consensus sequence is shown at the bottom of each column, with conserved amino acid motifs underlined. For full protein names, see Supplementary material

Structural or functional domain elements are generally formed by grouping of conserved amino acids. The 44 amino acids are very well conserved, with few similar amino acid residue substitutions not altering the overall property. For example, in cases where the consensus residue is an aromatic amino acid (Trp, Tyr, Phe) an aromatic substitution occurs in 87% of cases and non aromatic amino acids occurs only 13% of cases. For example, at position 339, which has conserved Tyr, has 6 aromatic amino acid substitutions (6 Phe) which are shown in figure 1. In position 561, which has conserved Asp residue is substituted with similar amino acid Glu in 8 sequences. In some cases, aliphatic character was completely conserved, such as position 220 Val is substituted by Ile 3 times.

Earlier, Ratnam, *et al*, in 1999 reported a model structure for FR- $\alpha$  using the crystal structure of Chicken riboflavin binding protein and have mentioned the ligand binding sites Ala<sup>49</sup>, Val<sup>104</sup>, and Glu<sup>166</sup> as important residues. [13] Position 49 specifically has small hydrophobic amino acids Ala/Leu. At position 49, Ala is highly conserved in all FR- $\alpha$  and FR- $\delta$  sequences, in case of FR- $\beta$  it is substituted by similar residue Leu and in case of binding proteins and FR- $\gamma$  it is Ala or Leu. Position 104 contains neutral, nonpolar amino acids with larger side chains-Phe, Val and Ile. Sequences of binding proteins, FR- $\alpha$ , FR- $\gamma$  and FR- $\delta$  have either Val or Ile, where as in FR- $\beta$  it is substituted by Phe. Position 166 contains neutral, nonpolar amino acid Gly or polar amino acids of various charges (positive or negative) aspartic acid, lysine, Gln, Glu and serine indicating less conservation in this position. Glu at 166 is conserved in 75% of FR- $\alpha$  sequence with remaining sequences having Gly or Lys as a substitution. FR- $\beta$  has Gly highly conserved except for *Equus caballus* sequence having Glu. FR- $\delta$  has Lys conserved except for *Canis familiaris* sequence where it is substituted by Lys. No such conservation can be concluded in FBP and FR- $\gamma$ . These substitutions are clearly shown in supplementary material. Present studies clearly indicate about the residues in binding pocket which lead to the specific affinity of subtypes of receptors towards various folate analogues.

We also compared the amino acid sequence of chicken RfBP with those of folate binding proteins and folate receptors. The Ligand binding site residue, Tyr-75 of RfBP is highly conserved in FBPs with few exceptions (His in four sequences and Cys in one sequence). Another ligand binding site, Trp-156 of RfBP is conserved in most of the FBP sequences except for one substitution Phe residue (Data not showed).

### Conserved protein motifs

A number of conserved amino acid motifs are shown at the bottom of Figure 1. Using PROSITE database we found out that 4 motifs are unique to folate binding proteins and the rest are observed in other protein sequences as well. The motif [LM]-L-[NS]-[VI]-C-M-

x(2)-[KR]-[HRY]-H-K (positions 217-228), C-x(3)-[TV]-S-x-[EAH]-[ALD]-[HT]-x-[DEA]-x-[SP]-x-[LS]-[YF]-x-F-[NST]-x(2)-H-C-[GS]-x-[ML]-x(3)-[CR] (positions 323-353), H-F-[IV]-Q-[DAN]-x-C-[LF]-[YHC]-E-C-S-P-N-L-G-P-W-[IF] (positions 370-388), E-D-C-x(2)-[WR]-W-x-[DA]-C-x(2)-[SY]-x-T-C-[KR]-x-[NDS] (positions 464-482) are uniquely present in folate receptors and folate binding proteins only. The other motifs such as, P-[GS]-[PQ]-E-[DG]-x-L-[HY] (positions 244-251), P-W-x(2)-[NRK]-[AS]-C (positions 309-315), [QRD]-x-[VAE]-x-[QSP]-x-[WGR]-x-E (positions 400-409), R-[FVI]-x(3)-P-L-C (positions 437-444), W-x(2)-[GS]-W-x-[WC] (positions 486-492), [YNHVF]-[FA]-P-[TGS]-[PGS]-[AKDTEV] (positions 523-528), [WC]-[STNDL]-[HNRVF]-[STD]-[YFN]-[KNE] positions 535-540, [YEA]-[SRGQ]-[RK]-[GNT]-[SQ]-G-[RQK]-[CLG]-[ILK]-[QD]-[MKH]-[WP]-F-[DE]-[PSAL]-[ATIVEF]-[QLEH]-[GSD]-N-P-N-[EV]-[EADV]-V-[AV]-[RKL]-[FLYH]-[YF]-[AL] (positions 548-576) (where x is any amino acid). This motif E-C-S-P-N-L-G-P-W (position 379-387) is 100% conserved in all the folate receptor sequences.

### Conserved cysteines residues

Cysteine residues form disulfide bridges that help to keep the molecule intact and to maintain the confirmation of elements of the active site. Monaco, in 1997, comparing amino acid sequences of chicken RfBP, bovine milk folate-binding protein and human folate-binding protein, identified 16 conserved cysteine residues that formed intra molecular disulfide bonds. [14] All analyzed 46 sequences in these substitution 16 cysteine residues. Out of 16, 10 cysteines are 100% aligned in our alignment positions 221, 315, 323, 346, 376, 380, 444, 466, 473, and 479. Remaining 6 cysteine residues are not fully conserved in our alignment (positions 307, 353, 499, 517, 531, and 555).

### Conclusion:

In our study 46 different FBP sequences were subjected to multiple sequence alignment to identify sequence homology and evolutionarily conserved residues, which found to be functionally important. Multiple sequence alignment has indicated that all 46 sequences have 44 highly conserved amino acids including 10 cysteines and 12 sequence motifs. The motifs obtained from multiple sequence alignment were compared with PROSITE database and we identified 4 motifs unique to folate binding protein sequences which further strengthened our belief that they are functionally important residues and highly conserved during evolution.

Two unique conserved motifs [LM]-L-[NS]-[VI]-C-M-x(2)-[KR]-[HRY]-H-K and C-x(3)-[TV]-S-x-[EAH]-[ALD]-[HT]-x-[DEA]-x-[SP]-x-[LS]-[YF]-x-F-[NST]-x(2)-H-C-[GS]-x-[ML]-x(3)-[CR] are predominantly helices and coils in most of the sequences. The other motif E-D-C-x(2)-[WR]-W-x-[DA]-C-x(2)-[SY]-x-T-C-[KR]-x-[NDS] form coil-helix-coil in most cases (FR- $\alpha$ ,

$\gamma$ , and  $\delta$ ) or only coil or helix-coil in case of beta. This conserved motif ECSPNLGPW is forming a coil between a strand and a helix. Most of these conserved motifs form coils are mostly not part of any secondary structures in all 46 sequences emphasizing its probable functional importance as loop regions frequently form binding sites, active sites. The predominance of helices along with coils may be due to its transmembrane location.

Ratnam *et al* in 1999 have reported functionally important residues in FR- $\alpha$  as Ala49, Val104, and Glu166 and in FR- $\beta$  as Leu49, Phe104 and Gly166. [13] Our data has shown that functionally important residues in FR- $\alpha$  and FR- $\beta$  are highly conserved in all organisms. In FR- $\alpha$  Ala in 49, Val/Ile in 104 similar amino acid substitution, Glu in 166, in case of FR- $\beta$  Leu in 49, Phe in 104 and Gly in 166. Our analytical data also supported the findings of similar functionally important residues in  $\delta$  as Ala49 and Val104. Similar amino acid substitution Leu/Ile was observed in three cases and Lys166 with one exception of Serine. This change of amino acid from Glu to Lys (166) in alpha and delta respectively may be the reason for their differences in function even though the other two residues Ala49 and Val104 are the same. This explains the probable difference in affinities of various receptors to folate and its analogues although it needs to be validated experimentally.

#### References:

[01] A. Brzezinska, *et al.*, *Acta Biochem Pol.*, 47: 735 (2000) [PMID: 11310973]

- [02] R. G. Andeson, *et al.*, *Science*, 255: 410 (1992) [PMID: 1310359]
- [03] R. S. Verma, *et al.*, *J Biol Chem.*, 67: 4119 (1992)
- [04] K. G. Rothberg, *et al.*, *J Cell Biol.*, 111: 2931 (1990)
- [05] F. Shen, *et al.*, *Biochem.*, 34: 5660 (1995) [PMID: 7727426]
- [06] A. Kamen, *et al.*, *J Biol Chem.*, 263: 13602 (1988)
- [07] H. Elnakat and M. Ratnam, *Adv. Drug Deliv Rev.*, 56: 1067 (2004)
- [08] M. Ratnam, *et al.*, *Biochem.*, 28: 8249 (1989) [PMID: 2605182]
- [09] P. C. Elwood, *J. Biol. Chem.*, 264: 14893 (1989) [PMID: 2768245]
- [10] H. Wang, *et al.*, *Nucleic Acids Res.*, 26: 2132 (1998) [PMID: 9547271]
- [11] W. Yan and M. Ratnam, *Biochem.*, 34: 14594 (1995) [PMID: 7578066]
- [12] O. Spiegelstein, *et al.*, *An International Journal of Genes, Genomes and Evol.*, 258: 117 (2000)
- [13] K. M. Maziarz, *et al.*, *J Biol Chem.*, 274: 11086 (1999) [PMID: 10196192]
- [14] L. H. Monaco, *EMBO Journal*, 16: 1475 (1997) [PMID: 9130692]
- [15] A. Benson, *et al.*, *Nucleic Acids Res.*, 30: 17 (2002)
- [16] J. D. Thompson, *et al.*, *Nucleic Acids Res.*, 22: 4673 (1994) [PMID: 7984417]
- [17] <http://www.expasy.ch/PROSITE/>

Edited by P. Kanguane

Citation: Ramamoorthy *et al.*, *Bioinformatics* 2(4): 157-162 (2007)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

### Supplementary material

Acc no	Description	Position 49	Position 104	Position 166
gi 18380986	Folr2 protein [Mus musculus]	L	F	G
gi 148684580	Folate receptor 2 (fetal), isoform CRA_b [Mus musculus]	L	F	G
gi 109462538	PREDICTED: similar to Folate receptor beta precursor [Rattus norvegicus]	L	F	G
gi 149719667	PREDICTED: similar to folate receptor 2 (fetal) [Equus caballus]	L	F	E
gi 62896637	Folate receptor 2 precursor variant [Homo sapiens]	L	F	G
gi 148277007	Folate receptor 2 precursor [Homo sapiens]	L	F	G
gi 114639250	PREDICTED: folate receptor 2 isoform 6 [Pan troglodytes]	L	F	G
gi 114639248	PREDICTED: similar to Folate receptor 2 isoform 7 [Pan troglodytes]	L	F	G
gi 109107807	PREDICTED: folate receptor 2 [Macaca mulatta]	L	F	G
gi 114639236	PREDICTED: similar to Folate receptor 2 isoform 3 [Pan troglodytes]	L	F	G
gi 73988197	PREDICTED: similar to Folate receptor beta precursor [Canis familiaris]	L	I	G
gi 115496686	Folate receptor 2 [Bos taurus]	L	F	G
gi 55742703	Folate binding protein [Sus scrofa]	L	I	G
gi 1169723	Folate receptor gamma precursor (FR-gamma) Homo sapiens	L	I	G
gi 114639226	PREDICTED: similar to folate receptor FRGAMMA 3 [Pan troglodytes]	L	I	G
gi 62898786	Folate receptor 1 precursor variant [Homo sapiens]	A	V	E
gi 49456991	FOLR1 [Homo sapiens]	A	V	E
gi 114639246	PREDICTED: folate receptor 1 isoform 1 [Pan troglodytes]	A	V	E
gi 114639240	PREDICTED: folate receptor 1 isoform 8 [Pan troglodytes]	A	V	E
gi 109107805	PREDICTED: similar to folate receptor 1 precursor isoform 2 [Macaca mulatta]	A	V	E
gi 114639238	PREDICTED: folate receptor 1 isoform 2 [Pan troglodytes]	A	V	G
gi 119907523	PREDICTED: similar to folate receptor alpha [Bos taurus]	A	I	E
gi 4928859	Membrane-bound folate binding protein [Sus scrofa]	A	I	Q
gi 73988199	PREDICTED: similar to folate receptor 1 (adult) [Canis familiaris]	A	I	E
gi 544338	Folate receptor alpha precursor [Mus musculus]	A	I	E
gi 19424178	Folate receptor 1 (adult) [Rattus norvegicus]	D	I	K
gi 149068707	Folate receptor 1 (adult), isoform CRA_b [Rattus norvegicus]	A	I	K
gi 5823002	Folate receptor alpha [Cricetulus griseus]	A	I	K
gi 149719669	PREDICTED: similar to folate receptor alpha [Equus caballus]	A	I	K
gi 149702958	PREDICTED: similar to folate receptor alpha [Equus caballus]	A	I	K
gi 110282963	Folate receptor alpha precursor (FR-alpha) (Milk folate-binding protein) Bovine	A	V	K
gi 47523188	Secreted folate binding protein [Sus scrofa]	A	I	K
gi 126327964	PREDICTED: similar to secreted folate binding protein [Monodelphis domestica]	A	V	N
gi 126327958	PREDICTED: similar to secreted folate binding protein [Monodelphis domestica]	A	V	N
gi 126327962	PREDICTED: similar to secreted folate binding protein [Monodelphis domestica]	A	V	N
gi 45382285	Folate receptor 1 (adult) [Gallus gallus]	A	I	K
gi 122937406	Similar to folate receptor 4 (delta) isoform 1 [Homo sapiens]	A	V	K
gi 114639852	PREDICTED: similar to Folate receptor 4 (delta) [Pan troglodytes]	A	V	K
gi 109108355	PREDICTED: similar to folate receptor 4 (delta) isoform 1 [Macaca mulatta]	A	V	K
gi 11527086	Folate receptor 3 [Mus musculus]	A	V	K
gi 20306583	Folate receptor 4 (delta) [Mus musculus]	A	V	K
gi 28827784	Folate receptor 4 (delta) isoform 2 [Mus musculus]	A	V	-
gi 148693064	Folate receptor 4 (delta), isoform CRA_a [Mus musculus]	A	V	-
gi 109484247	PREDICTED: similar to folate receptor 4 (delta) [Rattus orvegicus]	A	L	K
gi 109483088	PREDICTED: similar to folate receptor 4 (delta) isoform 1 [Rattus norvegicus]	A	L	K
gi 73987778	PREDICTED: similar to folate receptor 4 (delta) isoform 1 [Canis familiaris]	A	I	S

List of sequences used for multiple sequence alignment and three binding site residue and their convergence among all the folate receptor sequences