



**VOLGOGRAD
STATE
MEDICAL
UNIVERSITY**



**SECHENOV
UNIVERSITY**

Database queries for prediction and analysis of the spatial structure of the protein disulfide isomerase enzyme of the A4 family of humans and animals

R.V. Sushkov¹, A.V. Mochalov¹, Y.L. Orlov^{2, 3}, E.A Savina²

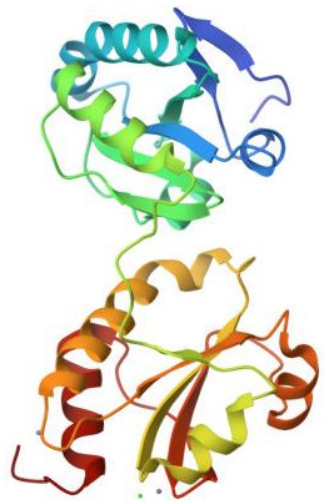
¹ Volgograd State Medical University, Volgograd, Russia

² Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

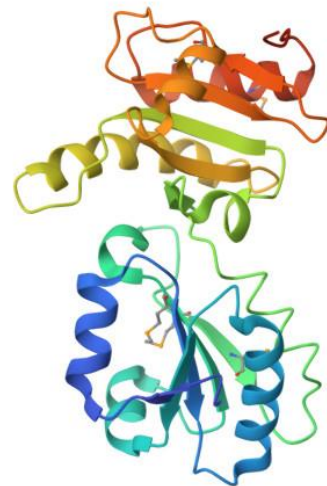
³ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

Aim: Protein disulfide isomerase of the A4 family (PDIA4) isomerase, catalyzes the formation of disulfide bonds and controls rearrangements of molecules in the EPR, which ensures their "flexibility", forms chaperone-mediated protein folding. Violation of the above functions leads to various degenerative changes in the nervous system, for example, Alzheimer's and Parkinson's diseases. The practical purpose of this work is to search for the most similar in spatial structure and functional activity of human and animal PDIA4 proteins, which can be used in medicine to create medicines based on its analogues, and for further research of the biochemical properties of biomolecules in the laboratory.

Materials and methods: Protein lists were compiled based on queries to the UniProtKB and BioGraph databases. Pairwise alignment of human and animal PDIA4 amino acid sequences was performed in the Clustal Omega program. The obtained probabilities of the difference between the two sequences were entered into an Excel spreadsheet. In parallel with the above, a pairwise comparison of amino acid sequences in the BLAST system was performed. The data were processed using rank estimates, on the basis of which a summary table was obtained, including three groups of proteins: the most similar to human protein, the least similar and the median. Next, using the RCSB site, 3D models of PDIA4 proteins were built. Amino acids subject to modifications in human and animal proteins were compared.



Protein disulfide isomerase of the A4 (Human)



Protein disulfide isomerase of the A4 (Rat)

| Animals studied | |
|------------------|---------------------------|
| Beaver | Greater horseshoe bat |
| Golden hamster | Kangaroo rat |
| Horse | Bovine |
| Ground squirrel | Bos indicus |
| Naked mole rat | Dog |
| Pig | Sheep |
| Black bear | Goat |
| Little brown bat | Chicken |
| Leopard | Duckbill |
| Cat | Mountain lion |
| Mustela | Gray short-tailed opossum |
| Weddell seal | American chameleon |
| Seal | Zebrafish |
| Rat | Gaboon caecilian |

```

sp|P13667|PDIA4_HUMAN MRPRKAFLLLLLLGLVQLLAVAGAEGPDEDSSNRENAIEDEEEEEEDDDEEDDLEVKE 60
sp|P38659|PDIA4_RAT MKLRKAWLLVLLALTLQLLAAASAEADAHEDASDSENPIEDDDDEE--EDEEEDDLEVKE 58
*:***:***:***.*.***.*.***. .**:*: ** ***::** :*:*:*****

sp|P13667|PDIA4_HUMAN ENGLVVLNDANFDNFVADKDTVLLFEYAPWCGHCKQFAPEYEKIANILKDKDPPPIVAKI 120
sp|P38659|PDIA4_RAT ENGVWVLNDENFDNFVADKDTVLLFEYAPWCGHCKQFAPEYEKIASTLKDNDPPPIAVAKI 118
**** **

sp|P13667|PDIA4_HUMAN DATSASVLASRFDVSGYPTIKILKKGQAVDYEGSRTQEEIVAKVREVSQPDWTPPPEVTL 180
sp|P38659|PDIA4_RAT DATSASMLASKFDVSGYPTIKILKKGQAVDYDGSRTQEEIVAKVREVSQPDWTPPPEVTL 178
*****:***:*****:*****:*****:*****

sp|P13667|PDIA4_HUMAN VLTKENFDEVVNDADIILVEFYAPWCGHCKKLAPEYEKAAKELSKRSPPIPLAKVDATAE 240
sp|P38659|PDIA4_RAT TLTKENFDVVNNDADIILVEFYAPWCGHCKKLAPEYEKAAKELSKRSPPIPLAKVDAT EQ 238
.*****:***:*****:*****:*****:*****:

sp|P13667|PDIA4_HUMAN TDLAKRFDVSGYPTLKIFRKG RPYDYGPREKYGIVDYMI EQSGPPSKEILTLKQVQEF 300
sp|P38659|PDIA4_RAT TDLAKRFDVSGYPTLKIFRKG RPFYDYGPREKYGIVDYMVEQSGPPSKEILTLKQVQEF 298
*****:***:*****:*****:*****:*****

sp|P13667|PDIA4_HUMAN KDGDDVIIIGVFKGESDPAYQQYQDAANLREDYKFHHTFSTEIAKFLKVSQGLVVMQP 360
sp|P38659|PDIA4_RAT KDGDDVILGVFGVGDGPGYLQYQDAANTLREDYKFHHTFSTEIAKFLKVSGLVLMQP 358
*****:***:***.*.***.* *****.*****:*****:*****:***

sp|P13667|PDIA4_HUMAN EKFSQSKYEPRSHMDVQGSTQDSAIKDFVLKYALPLVGHHRKVSNDAKRYTRRPLVVVYYS 420
sp|P38659|PDIA4_RAT EKFSQSKYEPRSHMDVQGSTESAIAKDYVVKHALPLVGHHRKVSNDAKRYTRRPLVVVYYS 418
***** *:*:*****: *****:***:*****.*****:*****

sp|P13667|PDIA4_HUMAN VDFSFDRYRAATQFWRSKVLEVAKDFPEYTFIAI DEEDYAGEVKDLGLSEGEDVNAAILD 480
sp|P38659|PDIA4_RAT VDFSFDRYRTATQFWRNKVLEVAKDFPEYTFIAI DEEDYATEVKDLGLSEGEDVNAAILD 478
*****:*****.*****:*****:***** *****

sp|P13667|PDIA4_HUMAN ESGKKFAMEPEEFDSDTLREFVTAFAKKGKLPVKSQPVKNNKGPVKVVVGGKTFDSIVM 540
sp|P38659|PDIA4_RAT ESGKKFAMEPEEFDSDALQEFVMAFKKGLKLPVKSQPVKNNKGPVVKVVVGGKTFDAIVM 538
*****:*****:*** *****:*****:*****:*****:*****:***

sp|P13667|PDIA4_HUMAN DPKKDVLIEFYAPWCGHCKQLEPVVNSLAKKYKGGKGLVIAKMDATANDVPSDRYKVEGF 600
sp|P38659|PDIA4_RAT DPKKDVLIEFYAPWCGHCKQLEPVVYSLGKKYKGGKDLVIAKMDATANDITNDRYKVEGF 598
*****:*****.*.*****.*****:***** *****

sp|P13667|PDIA4_HUMAN PTIYFAPSGDKKNPVKFEGGDRDLEHLSKFIEEHATKLSRTKEEL 645
sp|P38659|PDIA4_RAT PTIYFAPSGDKKNPIKFEGGDRDLEHLSKFIDEHATKRSRTKEEL 643
*****:*****:*****:***** *****

```

Pairwise comparison of human PDIA4 amino acid sequences with other organisms in the Clustal Omega program.

| Score | Expect | Method | Identities | Positives | Gaps |
|-----------------|--------|-----------------------------------------------------------------|--------------|--------------|-----------|
| 1029 bits(2661) | 0.0 | Compositional matrix adjust. | 505/647(78%) | 574/647(88%) | 8/647(1%) |
| Query 1 | 1 | MRPRKAFLLLLLLGLVQLLAVAGAEGPDEDSSNRENAIE--DEEEEEEDDDEEDDLEV | 58 | | |
| Sbjct 1 | 1 | MR R ++LLLLLGL+ QL + E RE+ + D +++EEE++D+++D+ EV | 54 | | |
| Query 59 | 59 | KEENGLVVLNDANFDNFVADKDTVLLFEYAPWCGHCKQFAPEYEKIANILKDKDPPPIVA | 118 | | |
| Sbjct 55 | 55 | KEENLVVLNDANFDFTADKDTVLLFEYAPWCGHCKQFAPEYEKIAKTLKENDPPPIVA | 114 | | |
| Query 119 | 119 | KIDATSASVLASRFDVSGYPTIKILKKGQAVDYEGSRTQEEIVAKVREVSQPDWTPPPEV | 178 | | |
| Sbjct 115 | 115 | KIDATAA+ LASRFDVSGYPTIKILKKGQ VDY+GSRT++ IVAKV+E+S P+WTPPPE | 174 | | |
| Query 179 | 179 | TLVLTKENFDEVVNDADIILVEFYAPWCGHCKKLAPEYEKAAKELSKRSPPIPLAKVDAT | 238 | | |
| Sbjct 175 | 175 | TLVLTQDNFDDVVKDADIILVEFYAPWCGHCKRLAPEYEKAAQELSKRTPIPLAKVDAT | 234 | | |
| Query 239 | 239 | AETDLAKRFDVSGYPTLKIFRKG RPYDYGPREKYGIVDYMI EQSGPPSKEILTLKQVQE | 298 | | |
| Sbjct 235 | 235 | AETELAKKFDVTGYPTLKIFRKG RPYDYGPREKYGIVDYMI EQAGPPSKQIQATKQVQE | 294 | | |
| Query 299 | 299 | FLKDGDDVIIIGVFKGESDPAYQQYQDAANLREDYKFHHTFSTEIAKFLKVSQGLVVM | 358 | | |
| Sbjct 295 | 295 | FLKDGDDVIIIGVFGSETDEYVQLYQEAANSLREDYKFHHTFSSEIAKLLKVSQGLVVM | 354 | | |
| Query 359 | 359 | QPEKFSQSKYEPRSHMDVQGSTQDSAIKDFVLKYALPLVGHHRKVSNDAKRYTRRPLVVVY | 418 | | |
| Sbjct 355 | 355 | QPEKFSQSK+EP+ +++D++ ST +S IK+ V+K+ALPLVGHHRK VSNDAKRY +RPLVVVY | 414 | | |
| Query 419 | 419 | YSVDFSFDRYRAATQFWRSKVLEVAKDFPEYTFIAI DEEDYAGEVKDLGLSEGEDVNAAI | 478 | | |
| Sbjct 415 | 415 | Y+VDFSFDRY ATQ+WR KVLEVAKDFPEY FA++DEEDY+ E+KDLGL ESGEDVN AI | 474 | | |
| Query 479 | 479 | LDESGKKFAMEPEEFDSDTLREFVTAFAKKGKLPVKSQPVKNNKGPVKVVVGGKTFDSI | 538 | | |
| Sbjct 475 | 475 | LDEGGKKYAMEPEEFDSDALRQFVLAFAKKGKLPVKSQPVKNNKGPVKVVVGGKTFDI | 534 | | |
| Query 539 | 539 | VMDPKDVLIEFYAPWCGHCKQLEPVVNSLAKKYKGGKGLVIAKMDATANDVPSDRYKVE | 598 | | |
| Sbjct 535 | 535 | VMDPK DVLIEFYAPWCGHCK+LEPVY L KKYK +K LVIKMDATANDV +D YKVE | 594 | | |
| Query 599 | 599 | GFPTIYFAPSGDKKNPVKFEGGDRDLEHLSKFIEEHATKLSRTKEEL 645 | | | |
| Sbjct 595 | 595 | GFPTIYFAPRDKKNPIKFEGGDRDLEHLSKFIEEHATKLSRTKEEL 641 | | | |

Pairwise comparison of human PDIA4 amino acid sequences with other organisms in the BLAST system.

Rank estimation of sequence similarity

| Организм | Clustal | BLAST | | | | | Clustal | BLAST | | | | | Средний ранг BLAST | Общий средний ранг |
|----------------|---------|-------------|-------------|----------|-------|-----------|---------|-------------|-------------|---------|-------|-----------|--------------------|--------------------|
| | p | Total score | Query cover | E value | Ident | Positives | Ранг p | Total score | Query cover | Ранги E | Ident | Positives | | |
| Бобр Beaver | 0,045 | 1155 | 100 | 0,00E+00 | 90,70 | 95 | 1 | 1 | 1 | 1 | 3 | 4 | 2,00 | 1,50 |
| Хомяк Golden | 0,049 | 1121 | 96 | 0,00E+00 | 90,18 | 96 | 2 | 2 | 14 | 1 | 5 | 1 | 4,60 | 3,30 |
| Лошадь Horse | 0,050 | 1118 | 100 | 0,00E+00 | 90,08 | 95 | 4 | 3 | 1 | 1 | 8 | 4 | 3,40 | 3,70 |
| Суслик Ground | 0,049 | 1113 | 90 | 0,00E+00 | 91,44 | 96 | 3 | 7 | 23 | 1 | 1 | 1 | 6,60 | 4,80 |
| Голые землекоп | 0,052 | 1118 | 100 | 0,00E+00 | 88,99 | 95 | 6 | 3 | 1 | 1 | 12 | 4 | 4,20 | 5,10 |
| Свинья Pig | 0,051 | 1117 | 96 | 0,00E+00 | 90,18 | 95 | 5 | 5 | 14 | 1 | 5 | 4 | 5,80 | 5,40 |
| Медведь Ame | 0,053 | 1115 | 96 | 0,00E+00 | 90,16 | 95 | 7 | 6 | 14 | 1 | 7 | 4 | 6,40 | 6,70 |
| Малая бурьян | 0,055 | 1108 | 90 | 0,00E+00 | 90,92 | 96 | 8 | 8 | 23 | 1 | 2 | 1 | 7,00 | 7,50 |
| Леопард Leop | 0,055 | 1106 | 100 | 0,00E+00 | 88,68 | 95 | 9 | 10 | 1 | 1 | 15 | 4 | 6,20 | 7,60 |
| Кот Cat | 0,056 | 1101 | 100 | 0,00E+00 | 88,70 | 95 | 10 | 13 | 1 | 1 | 14 | 4 | 6,60 | 8,30 |
| Хорек Mustela | 0,057 | 1106 | 96 | 0,00E+00 | 89,52 | 95 | 11 | 10 | 14 | 1 | 9 | 4 | 7,60 | 9,30 |
| Тюлень Wedd | 0,057 | 1107 | 96 | 0,00E+00 | 89,35 | 95 | 11 | 9 | 14 | 1 | 11 | 4 | 7,80 | 9,40 |
| Морской котик | 0,057 | 1101 | 96 | 0,00E+00 | 88,87 | 95 | 13 | 13 | 14 | 1 | 13 | 4 | 9,00 | 11,00 |
| Крыса Rat | 0,058 | 1106 | 90 | 0,00E+00 | 90,58 | 95 | 14 | 10 | 23 | 1 | 4 | 4 | 8,40 | 11,20 |
| Большой под | 0,062 | 1088 | 90 | 0,00E+00 | 89,37 | 95 | 15 | 17 | 23 | 1 | 10 | 4 | 11,00 | 13,00 |
| Кенгуровые п | 0,062 | 1091 | 95 | 0,00E+00 | 88,64 | 95 | 15 | 15 | 20 | 1 | 16 | 4 | 11,20 | 13,10 |
| Бык Bovin | 0,064 | 1081 | 100 | 0,00E+00 | 86,67 | 94 | 17 | 18 | 1 | 1 | 17 | 17 | 10,80 | 13,90 |
| Зебу Bos indi | 0,065 | 1090 | 100 | 0,00E+00 | 86,36 | 93 | 18 | 16 | 1 | 1 | 18 | 19 | 11,00 | 14,50 |
| Собака Dog | 0,073 | 1072 | 100 | 0,00E+00 | 85,50 | 93 | 19 | 19 | 1 | 1 | 19 | 19 | 11,80 | 15,40 |
| Овца Sheep | 0,074 | 1069 | 100 | 0,00E+00 | 85,12 | 93 | 20 | 20 | 1 | 1 | 20 | 19 | 12,20 | 16,10 |
| Коза Goat | 0,074 | 1069 | 100 | 0,00E+00 | 85,12 | 93 | 20 | 20 | 1 | 1 | 20 | 19 | 12,20 | 16,10 |
| Курица Chick | 0,107 | 1029 | 100 | 0,00E+00 | 78,05 | 88 | 25 | 23 | 1 | 1 | 25 | 25 | 15,00 | 20,00 |
| Утка Duck | 0,084 | 1064 | 90 | 0,00E+00 | 68,47 | 94 | 22 | 22 | 23 | 1 | 28 | 17 | 18,20 | 20,10 |
| Горный лев M | 0,097 | 957 | 99 | 0,00E+00 | 81,00 | 87 | 23 | 28 | 12 | 1 | 23 | 26 | 18,00 | 20,50 |
| Опоссум Gray | 0,105 | 1014 | 91 | 0,00E+00 | 81,15 | 90 | 24 | 24 | 21 | 1 | 22 | 23 | 18,20 | 21,10 |
| Хамелеон Ame | 0,115 | 1010 | 91 | 0,00E+00 | 80,65 | 89 | 26 | 25 | 21 | 1 | 24 | 24 | 19,00 | 22,50 |
| Рыба-зебра Z | 0,136 | 960 | 99 | 0,00E+00 | 72,34 | 84 | 27 | 27 | 12 | 1 | 27 | 28 | 19,00 | 23,00 |
| Червяга Gabe | 0,138 | 965 | 90 | 0,00E+00 | 75,90 | 87 | 28 | 26 | 23 | 1 | 26 | 26 | 20,40 | 24,20 |

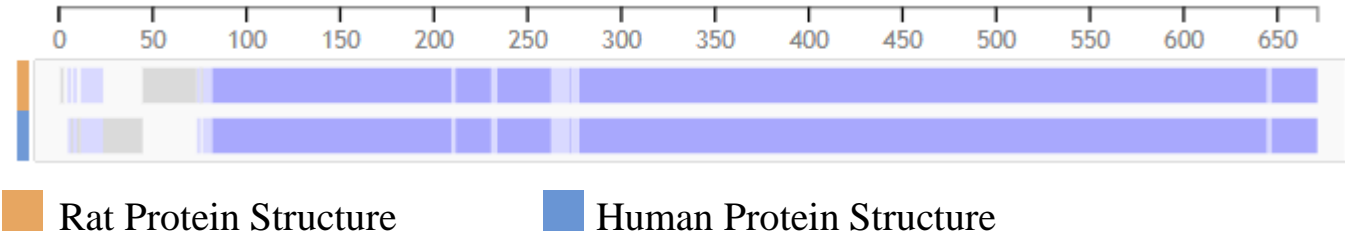
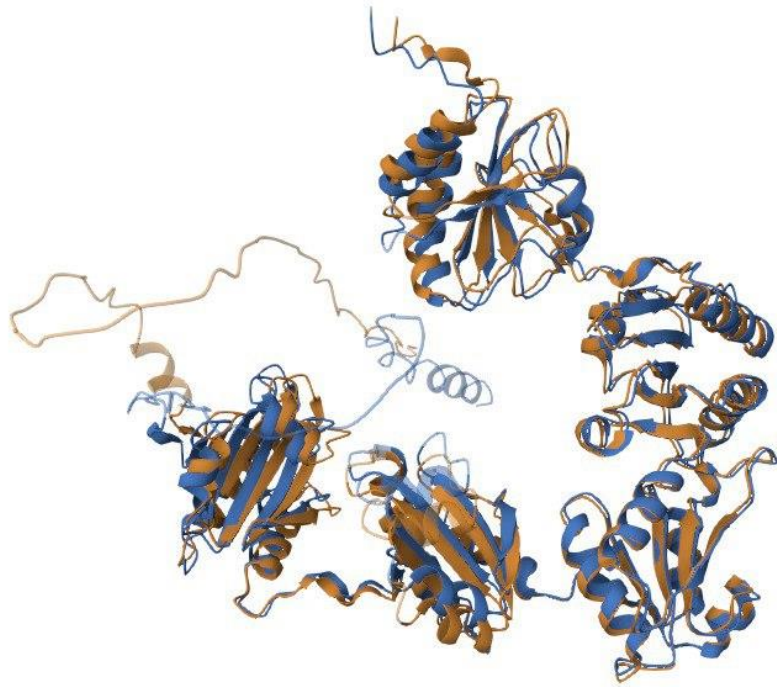
As a result of the pairwise comparisons, we obtain a large amount of data, including: Clustal P; Total score; Query cover; E-value; Indent; Positives.

We process the data obtained using rank estimates. As a result, we get the following table.

Organisms with the greatest similarity to the PDIA4 protein are colored green: Beaver, Golden hamster, Horse, Ground squirrel, Naked mole rat, Pig, Black bear, Little brown bat.

The organisms that are least similar to the human protein PDIA4 are highlighted in pink.

Yellow is the median, which is also suitable for further sequence analysis.



A comparison of the 3D structures of the human and rat PDIA4 protein on the RCSB website yielded a result of 0.88 TM-score, a high percentage of correlation.

Conclusion:

- As a result of the alignment of amino acid sequences in the Clustal Omega and BLAST system, the following was revealed: the hamster protein is most similar to the human variant of PDIA4; a hamster or rat should be used for in vivo studies, since they are a standard laboratory animal and have a high level of correlation; to obtain a pure enzyme and conduct in vitro studies it is more rational to consider a pig, due to the availability of its tissues;
- A comparison of the 3D structures of the human and rat PDIA4 protein on the RCSB website yielded a result of 0.88 TM-score, a high percentage of correlation.