DOI: 10.1002/cmdc.200700154

On the use of PIPSA to Guide Target-Selective Drug Design

Stefan Henrich,* Stefan Richter, and Rebecca C. Wade*[a]

In structure-based drug design, it is important to design compounds to bind specifically and selectively to their macromolecular target receptor(s). Binding to other macromolecules that are similar to the target may result in adverse side effects and should be avoided. The decision as to which macromolecular target and which region of the target a drug should bind to should therefore include consideration of the binding properties of related macromolecules. Herein, we show how PIPSA (protein interaction property similarity analysis)[1–3] can aid in surveying the interaction properties of structurally-related macromolecules before embarking on detailed design towards a chosen target site. This is illustrated by application of PIPSA to dihydrofolate reductase (DHFR; EC 1.5.1.3).

DHFR is an essential and conserved enzyme in many species. It takes part in folate metabolism and is important for thymidine synthesis. It binds the cofactor NADPH and converts dihydrofolate (DHF) into tetrahydrofolate (THF). As a result of its central role, DHFR is an important drug target and inhibitors, such as methotrexate (MTX), trimethoprim, and pyrimethamine, are used against cancer, bacterial, and parasitic diseases, respectively.[4] A challenge in the application of DHFR inhibitors as antibiotics is the occurrence of side effects. These may arise from the binding of the compounds to human DHFR. On the other hand, it may be advantageous for antibiotics to have a broad spectrum of activity and to bind to several microbial DHFRs. By way of example, we consider herein the selective targeting of compounds to *Candida albicans* DHFR. Such compounds would be particularly useful to treat common opportunistic infections in immunocompromised patients.[5] Although known clinical drugs against DHFR show weak activity against *C. albicans*,[5] potent, selective inhibitors of *C. albicans* DHFR have been reported.[6,7] Their selectivity for *C. albicans* versus human DHFR has been ascribed in part to differences in protein–ligand hydrogen bonding. Such differences can be detected by analysis of the protein electrostatic potentials.

PIPSA permits quantification of the similarity in the interaction properties of homologous proteins and has been applied to a variety of protein types.[3,8–11] PIPSA is available as standalone software[12] but has recently been made available online in the SYCAMORE webserver.[13] In the latter case, it is combined in a workflow with automated protein homology model building and electrostatic potential calculation. Herein, we demonstrate use of the online PIPSA workflow for DHFRs from different species.

[a] *Dr. S. Henrich, Dr. S. Richter, Dr. R. C. Wade*
EML Research GmbH
Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg (Germany)
Fax: (+49) 6221-533-298
E-mail: stefan.henrich@eml-r.villa-bosch.de
rebecca.wade@eml-r.villa-bosch.de

In the first step of PIPSA, one crystal structure of DHFR (human DHFR, Swiss-Prot identifier P00374) was chosen as a template for homology modeling (Figure 1). All related DHFR

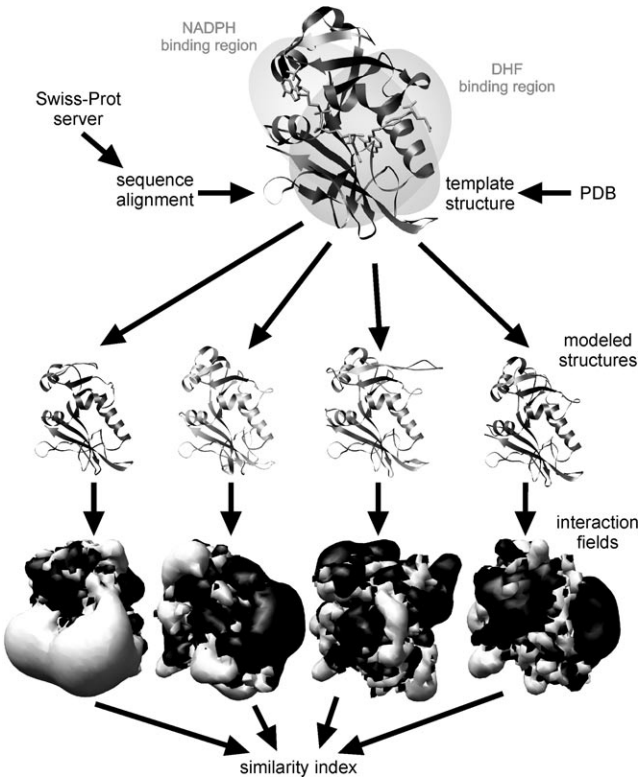

Figure 1. Schematic overview of the PIPSA workflow of sequence alignment, automatic homology modeling, and computation of molecular interaction fields and similarity indices (SI).

sequences for which kinetic data are available either in BRENDA[14] or SABIO-RK[15] were then retrieved from Swiss-Prot[16] and aligned to the template using CLUSTALW[17] and T-COFFEE.[18] Based on the alignment, homology models were generated automatically with Modeller.[19] For all protein models that passed structure validation checks with Modeller and WHATIF,[20] electrostatic potentials were calculated by solution of the linearized Poisson–Boltzmann equation using UHBD[21] on a grid of 1 Å spacing. For each pair of proteins in the dataset, the potentials at corresponding grid points within a region ("skin") of 4 Å thickness extending from 3 Å away from the protein surface were compared. Hodgkin similarity indices (SI)[22] were calculated for each protein pair. The SI values lie between 1 and −1, where SI = 1 indicates identical, SI = 0 uncorrelated, and SI = −1 anticorrelated electrostatic potentials. The calculated relationships between the protein electrostatic potentials were displayed either as a color-coded matrix of SI values using JMP6 (SAS Institute, USA) or as a dendro-

gram generated by R[23] (Figure 2). In addition to comparing the whole protein skins, SI values were computed for focused regions centered on DHF or NADPH (Figure 1).

To check the robustness of the protein models and the SI calculation, six crystal structures of phylogenetically distinct DHFR enzymes (PDB codes: 1u72 (*Homo sapiens*; animalia), 1rh3 (*Escherichia coli*; bacteria/Gram-negative), 1aeo (*C. albicans*; fungi/ascomycota), 1dyr (*Pneumocystis carinii*; fungi/ascomycota), 3dfr (*Lactobacillus casei*; bacteria/Gram-positive), and 1j3i (*Plasmodium falciparum*; alveolata/apicomplexa) were used as template structures in independent runs. Although, the alignments of the retrieved sequences to the different tem-

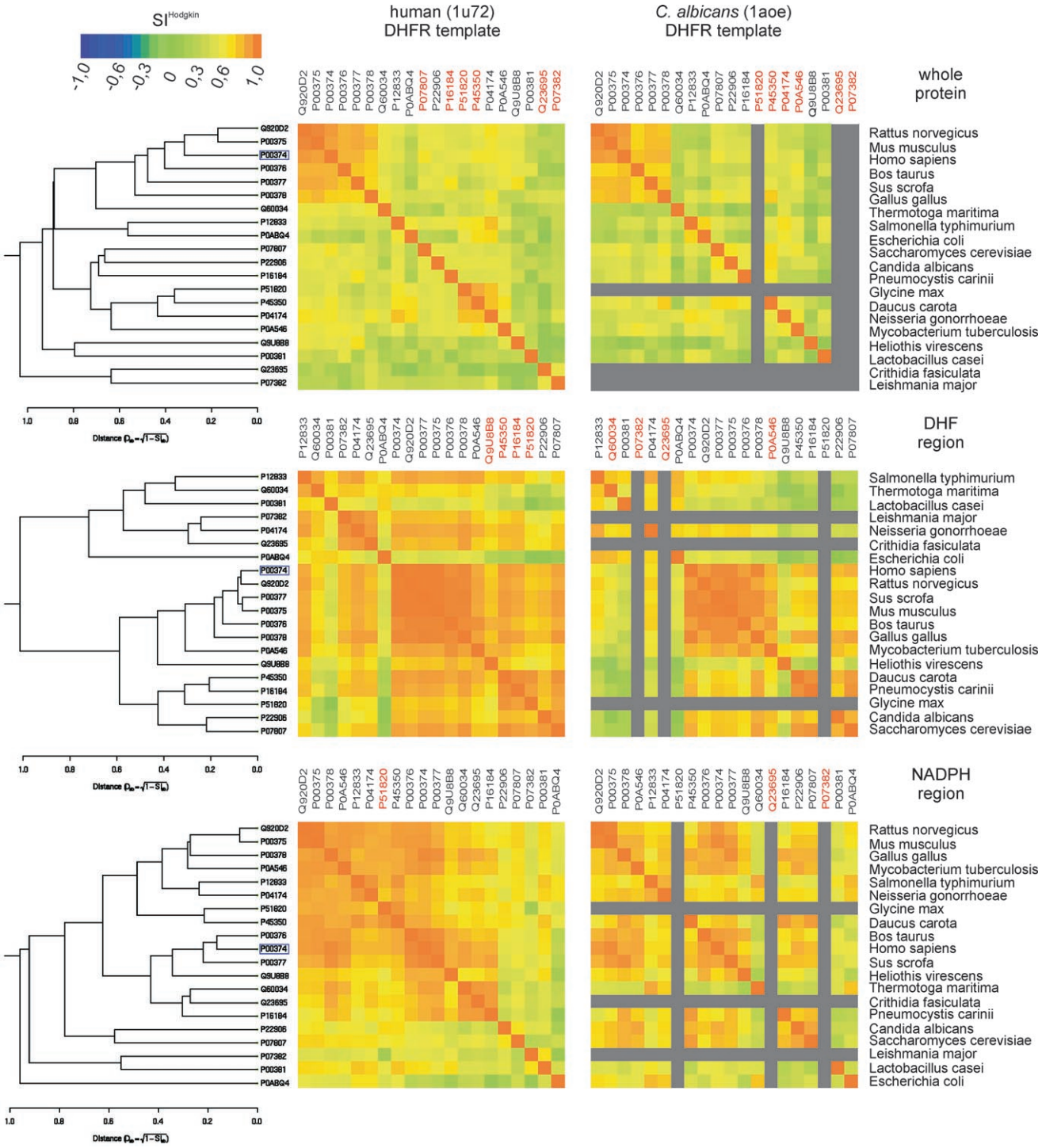

Figure 2. SI values for DHFR structures modeled on a human template structure (1u72, left color maps) are shown as dendrograms and color maps: whole protein (top), DHF (middle), and NADPH (bottom) binding regions. The right-hand color maps show SI values for proteins modeled on the *C. albicans* template structure (1aoe). The Swiss-Prot identifiers and corresponding species names are given as matrix labels. Models with no computed SI values and those with partly incorrectly modeled structures are shown with gray rows and columns or marked in red, respectively.

plates were reasonable, the modeled structures showed variations in the backbone and side-chain positions. For example, in some models, the side chain of an arginine in the active site pointed in opposite directions (see Figure 3). Furthermore, due to insertions or deletions, some loops were modeled without a corresponding template sequence. Therefore, some of the structures did not pass the WHATIF check and no SI values were calculated in the automated workflow (gray rows and columns in Figure 2). A manual inspection revealed further incorrect parts, although most models showed a reasonable fold in the focus regions. The problematic structures were still included in the PIPSA analysis but are marked in Figure 2 with a red Swiss-Prot identifier.

The three distance matrices computed using the human DHFR structure as the template for protein modeling were clustered and the species were arranged as shown in Figure 2 (left) (for simplicity, subtypes of *E. coli* DHFR, often MTX-resistant, were removed). The SI values obtained using the human (left) and *C. albicans* (right) DHFR structures as modeling templates are shown as color maps in Figure 2. The results are shown for the whole protein (top), the focused DHF (middle), and NADPH (lower) binding regions. Despite problems in modeling loop regions and side-chain positions in some structures (see Figure 3), the SI matrices derived for the three regions using the human and *C. albicans* DHFR templates are quite similar and comparable to those based on the *E. coli*, *L. casei*, *P. carinii*, and *P. falciparum* DHFR templates (data not shown). The color maps demonstrate the overall robustness of the analysis towards the chosen template structures but also highlight that, in some cases, not every sequence can be modeled onto every structural template satisfactorily. It is recommended to choose a suitable template closely related to the sequences of interest and to check the modeled proteins manually.

Dividing the DHFRs of different species into four groups according to the dendrograms suggests that, for the DHF binding region, phylogenetically related species occur in the same clusters (kinetoplasida/bacteria; *E. coli*; animalia/actinobacteria; plants/fungi). Clusters for the whole protein and the NADPH binding region are more heterogeneous and, in the latter case, not restricted to their evolutionary relationship (Figure 2 and Table 1). However, in all three cases, the fungi DHFRs form, either alone or together with the plant DHFRs, a distinct group. *C. albicans* is always clustered in a different group from human DHFR. It is interesting that the known selective *C. albicans* inhibitors bind in the DHF site and extend to the NADPH site which may allow them to exploit differences in the molecular interaction fields of both sites.

To compare the PIPSA results with experimental data on ligand binding, K_M values for the cofactor NADPH and the substrate DHF were collected from the literature. The experimental values for the proteins vary indicating that they are highly dependent on assay conditions (for example, K_M values for DHF and NADPH in human DHFR are reported as 0.032 to 2.8 and 0.16 to 9.1 μM, respectively). Nevertheless, the most reliable K_M values determined in low-salt concentration enzyme assays around pH 7 were taken and are given in Table 1 and plotted against each other in Figure 4 (right). After marking according to the four clusters of the PIPSA analysis for the whole protein (Table 1), a relation is revealed between the K_M values for DHF and their PIPSA clustering. Those in cluster W1 (Chordata, Thermotogae, and Protobacteria) have lower K_M values and those in cluster W2 (Ascomycota, Magnoliophyta, Proteobacteria, and Actinobacteria) have higher K_M values. The K_M values of cluster W4 (Euglenozoa) cannot be distinguished from those of cluster W1. Focusing only on the DHF substrate-binding region (symbols in Figure 4), a separation of the K_M values of DHF into a lower (cluster D1; kinetoplasida/bacteria) and a higher range (cluster D4; plants/fungi) was revealed. PIPSA analysis did not reveal a relation between K_M values for NADPH and the electrostatic potential on the whole protein or focused on the cofactor region. This indicates that other properties underlie the differences in K_M values for NADPH.

Most of the reported QSAR models for DHFR are based only on hydrophobicity, dipolar interactions and molar refraction (reviewed in [24]), but the necessity to quantify electrostatic potentials and their importance in ligand binding in DHFR has been pointed out.[25–27] A direct comparison of ligand-based QSAR and receptor-based PIPSA analysis is not easy, because in the former, the active site of

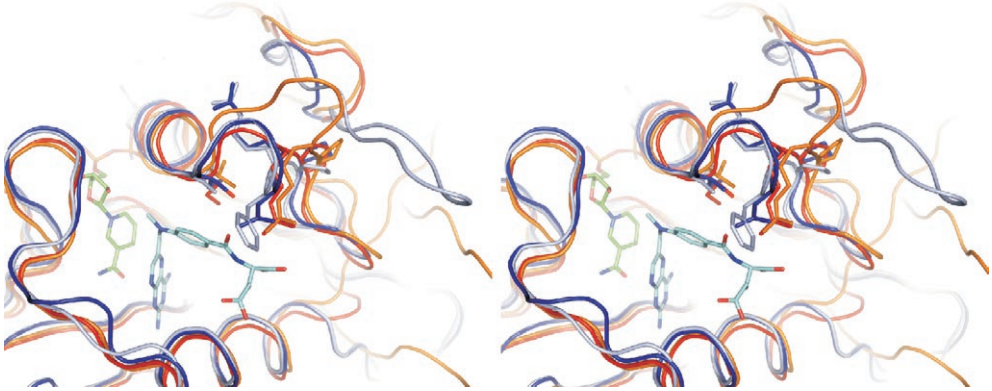

Figure 3. Stereo illustration of structural modeling problems that may affect PIPSA. Superposition of the crystal structures, 1u72 (human; blue) and 1rh3 (*E. coli*; red), with the 1u72-based (light blue) and 1rh3-based (orange) modeled structures of *Leishmania major* DHFR (P07382). The cofactor, NADPH (green), and the inhibitor, methotrexate (cyan), from the human DHFR structure, 1u72, are shown colored by atom type. Depending on the template used, the arginine side chain of the models points either towards (1rh3-based, superimposing on *E. coli* Arg 52) or away (1u72-based, superimposing on human Arg 65) from the active site because of a three residue-deletion in the *E. coli* template compared to the human template. For the latter model, instead of the arginine, Phe 69 points inwards superimposing on human Asn 62. Although the structure of the modeled turn differs, the conserved proline and isoleucine/valine flanking the arginine have similar positions. Additionally, for the 1u72-based model, a loop is shown in the upper right corner that was modeled without a corresponding template sequence.

Table 1. Overiew of PIPSA results given with published K_M values for NADPH and DHF for DHFRs from various species.[a]

species	phylum	clustering whole protein	DHF region	NADPH region	K_M [μM] DHF	NADPH	lit.
Rattus norvegicus	Chordata	W1	D3	N1	1.8	2.4	[28]
Mus musculus	Chordata	W1	D3	N1	0.30	1.36	[29]
Homo sapiens	Chordata	W1	D3	N1	0.11	2.5	[30]
Bos taurus	Chordata	W1	D3	N1	2.3	33	[31]
Sus scrofa	Chordata	W1	D3	N1	0.74	3.22	[29]
Gallus gallus	Chordata	W1	D3	N1	0.15	1.8	[32]
Thermotoga maritima	Thermotogae	W1	D1	N1	0.3	4.0	[33]
Salmonella typhimurium	Proteobacteria	W1	D1	N1	0.4	n. d.	[34]
Escherichia coli	Proteobacteria	W1	D2	N4	0.27	1.05	[35]
Saccharomyces cerevisiae	Ascomycota	W2	D4	N2	13	45	[36]
Candida albicans	Ascomycota	W2	D4	N2	2.7	3.6	[5]
Pneumocystis carinii	Ascomycota	W2	D4	N1	2.3	3.0	[37]
Glycine max	Magnoliophyta	W2	D4	N1	35	415	[36]
Daucus carota	Magnoliophyta	W2	D4	N1	3.7	2.2	[38]
Neisseria gonorrhoeae	Proteobacteria	W2	D1	N1	2.6	10	[39]
Mycobacterium tuberculosis	Actinobacteria	W2	D3	N1	4.5	4.2	[40]
Heliothis virescens	Arthropoda	W3	D3	N1	7.5	6.7	[41]
Lactobacillus casei	Firmicutes	W3	D1	N3	0.36	0.78	[42]
Crithidia fasciculata	Euglenozoa	W4	D1	N1	1.1	2.7	[43]
Leishmania major	Euglenozoa	W4	D1	N3	1.3	0.9	[44]

[a] The clustering gives four clusters based on the PIPSA analysis with the human template structure (see Figures 2 and 4).

DHFRs from a few species were analyzed using a series of similar ligands binding in a similar pose, and, in the latter, a comparison of electrostatic potentials covering a larger region of the protein is done. Furthermore, the K_i values of prominent inhibitors are not available for the DHFRs of all the species investigated and often there is a large inconsistency in the published data. Nevertheless, without prior knowledge, PIPSA can help to identify residues that are important for selectivity between different species. For example, DHFRs from species that form the conserved salt bridge to the nitrogens of DHF by glutamate (human Glu 30) are separated from those using aspartate (*E. coli* Asp 27) (Figure 2 and Figure 4). Furthermore, an interesting clustering can be found

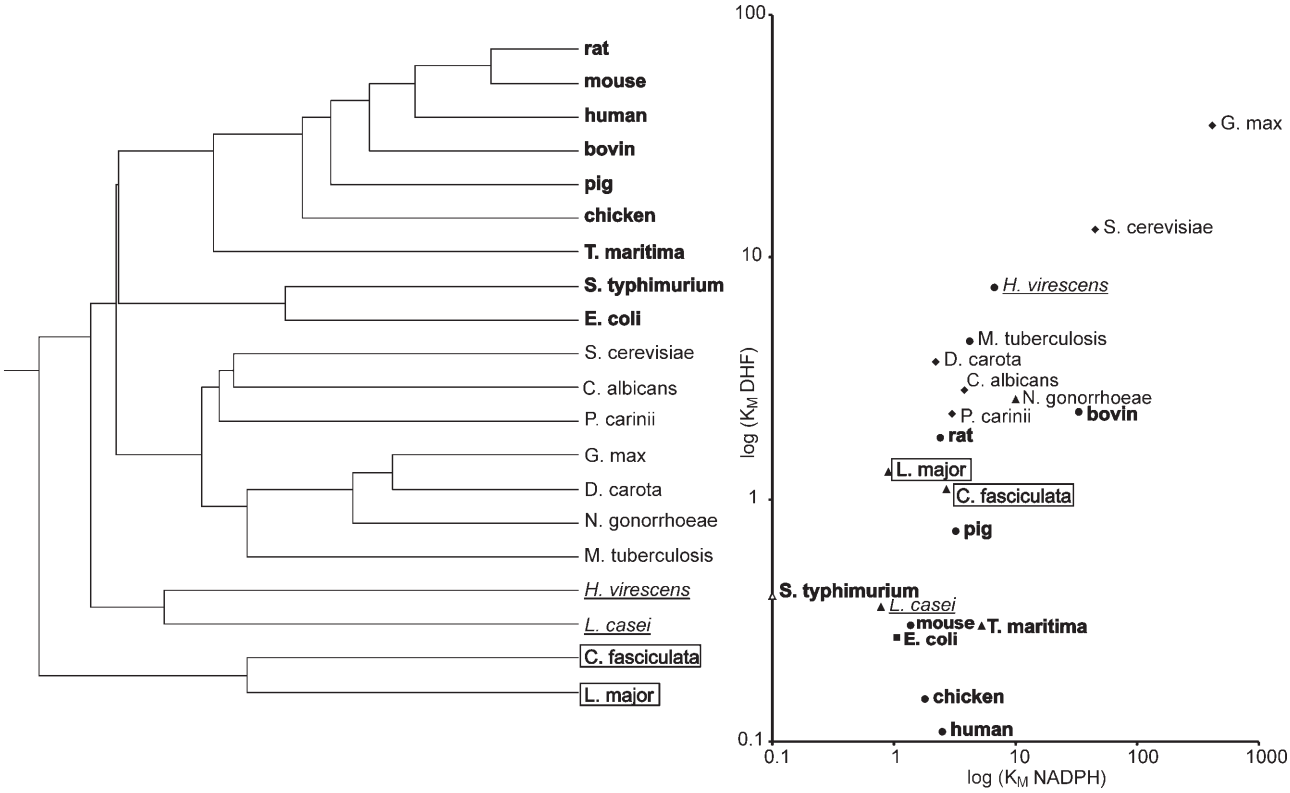

Figure 4. Dendrogram derived from PIPSA analysis for the whole DHFR protein based on the human template (left) with the four most distinct clusters marked separately (W1: bold, W2: standard font, W3: italic-underlined, W4: boxed). K_M values taken from the literature for the substrate DHF and the cofactor NADPH are plotted on a logarithmic scale against each other (right). The data points are marked as in the dendrogram of the left panel, whereas the symbols refer to the clustering based on PIPSA analysis focused on the DHF region (see Table 1 and middle panel of Figure 2; D1: triangles, D2: squares, D3: circles, D4: diamonds). For the DHFR from *S. typhimurium*, no K_M value for NADPH could be found.

for Asn64 (human DHFR), which is conserved in chordata and is located at the entrance of the active site forming hydrogen bonds, with, for example, MTX. The other members of cluster W1 show a different loop conformation in the crystal structures resulting in a replacement of Asn64 by Arg52 (*E. coli*), enabling them to also form hydrogen bonds to MTX (Figure 3). Even with problems of model building, they were grouped together in cluster W1 and were separated from all fungi DHFRs, where Asn64 is replaced by Phe and not able to form a hydrogen bond. This in agreement with kinetic data[5] and probably explains the strong binding of MTX to human and *E. coli* DHFRs as well as the weak binding to *C. albicans* DHFR.

In summary, we have presented how PIPSA can be used together with automatic homology modeling to quantify electrostatic similarities and differences between homologous proteins from different species. This can help to detect structural and functional relationships between the proteins, which are important to recognize to avoid side effects in drug design projects.

Acknowledgements

This research was supported by the BMBF (Systems Biology Initiative HepatoSys Grant numbers 0313076 and 0313078C), the EU STREP project LIGHTS (37852), the German-Italian project (MIUR-I04C0CGCE), and the Klaus Tschira Foundation. We thank Nils Semmelrock for assistance with the dendrograms.

Keywords: dihydrofolate reductase · drug selectivity · electrostatic interactions · homology modeling · molecular interaction field

[1] N. Blomberg, R. R. Gabdoulline, M. Nilges, R. C. Wade, *Proteins* **1999**, *37*, 379–387.
[2] R. C. Wade, R. R. Gabdoulline, F. De Rienzo, *Int. J. Quantum Chem.* **2001**, *83*, 122–127.
[3] P. J. Winn, M. Zahran, J. N. Battey, Y. Zhou, R. C. Wade, A. Banerjee, *Front. Biosci.* **2007**, *12*, 3419–3430.
[4] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, *Nucleic Acids Res.* **2006**, *34*, D668–672.
[5] D. P. Baccanari, R. L. Tansik, S. S. Joyner, M. E. Fling, P. L. Smith, J. H. Freisheim, *J. Biol. Chem.* **1989**, *264*, 1100–1107.
[6] J. H. Chan, J. S. Hong, L. F. Kuyper, D. P. Baccanari, S. S. Joyner, R. L. Tansik, C. M. Boytos, S. K. Rudolph, *J. Med. Chem.* **1995**, *38*, 3608–3616.
[7] M. Whitlow, A. J. Howard, D. Stewart, K. D. Hardman, J. H. Chan, D. P. Baccanari, R. L. Tansik, J. S. Hong, L. F. Kuyper, *J. Med. Chem.* **2001**, *44*, 2928–2932.
[8] F. De Rienzo, R. R. Gabdoulline, M. C. Menziani, P. G. De Benedetti, R. C. Wade, *Biophys. J.* **2001**, *81*, 3090–3104.
[9] F. De Rienzo, R. R. Gabdoulline, M. C. Menziani, R. C. Wade, *Protein Sci.* **2000**, *9*, 1439–1454.
[10] K. Schleinkofer, U. Wiedemann, L. Otte, T. Wang, G. Krause, H. Oschkinat, R. C. Wade, *J. Mol. Biol.* **2004**, *344*, 865–881.
[11] P. J. Winn, T. L. Religa, J. N. Battey, A. Banerjee, R. C. Wade, *Structure* **2004**, *12*, 1563–1574.
[12] http://projects.villa-bosch.de/mcmsoft/pipsa/2.0/.
[13] http://sycamore.eml.org.
[14] http://www.brenda-enzymes.info.
[15] U. Wittig, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, S. Anstein, J. Saric, I. Rojas in *Data Integration in the Life Sciences, Vol. 4075*, Springer, Berlin, **2006**, pp. 94–103.
[16] http://www.expasy.org.
[17] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson, *Nucleic Acids Res.* **2003**, *31*, 3497–3500.
[18] C. Notredame, D. G. Higgins, J. Heringa, *J. Mol. Biol.* **2000**, *302*, 205–217.
[19] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779–815.
[20] G. Vriend, *J. Mol. Graph.* **1990**, *8*, 52–56, 29.
[21] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, J. A. McCammon, *Comput. Phys. Commun.* **1995**, *91*, 57–95.
[22] E. E. Hodgkin, W. G. Richards, *Int. J. Quantum Chem. Quantum Biol. Symp.* **1987**, *32*, 105–110.
[23] R Development Core Team, R: A language and environment for statistical computing, 2007 http://www.R-project.org.
[24] C. D. Selassie in *Burger's Medicinal Chemistry and Drug Discovery, Vol. 1* (Ed.: D. J. Abraham), Wiley, New York, **2003**, p. 1.
[25] J. J. Sutherland, D. F. Weaver, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 309–331.
[26] A. Gangjee, X. Lin, *J. Med. Chem.* **2005**, *48*, 1448–1469.
[27] C. D. Selassie, Z. Fang, R. Li, C. Hansch, G. Debnath, T. Klein, R. Langridge, B. T. Kaufman, *J. Med. Chem.* **1989**, *32*, 1895–1905.
[28] H. Herken, R. Timmler, *Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmakol.* **1965**, *250*, 293–301.
[29] S. L. Smith, P. Patrick, D. Stone, A. W. Phillips, J. J. Burchall, *J. Biol. Chem.* **1979**, *254*, 11475–11484.
[30] J. Jarabak, N. R. Bachur, *Arch. Biochem. Biophys.* **1971**, *142*, 417–425.
[31] D. M. Greenberg, B.-D. Tam, E. Jenny, B. Payes, *Biochim. Biophys. Acta Enzymol. Biol. Oxid.* **1966**, *122*, 423–435.
[32] B. T. Kaufman, R. C. Gardiner, B. G. Tresvant, *J. Biol. Chem.* **1966**, *241*, 1319–1328.
[33] V. Wilquet, J. A. Gaspar, M. van de Lande, M. Van de Casteele, C. Legrain, E. M. Meiering, N. Glansdorff, *Eur. J. Biochem.* **1998**, *255*, 628–637.
[34] S. S. Joyner, M. E. Fling, D. Stone, D. P. Baccanari, *J. Biol. Chem.* **1984**, *259*, 5851–5856.
[35] M. H. Penner, C. Frieden, *J. Biol. Chem.* **1987**, *262*, 15908–15914.
[36] V. A. Reddy, N. A. Rao, *Arch. Biochem. Biophys.* **1976**, *174*, 675–683.
[37] C. J. Delves, S. P. Ballantine, R. L. Tansik, D. P. Baccanari, D. K. Stammers, *Protein Expression Purif.* **1993**, *4*, 16–23.
[38] D. Albani, B. Parisi, D. Carbonera, R. Cella, *Plant Mol. Biol.* **1985**, *5*, 363–372.
[39] D. P. Baccanari, R. L. Tansik, S. J. Paterson, D. Stone, *J. Biol. Chem.* **1984**, *259*, 12291–12298.
[40] E. L. White, L. J. Ross, A. Cunningham, V. Escuyer, *FEMS Microbiol. Lett.* **2004**, *232*, 101–105.
[41] V. K. Walker, M. G. Tyshenko, M. J. Kuiper, R. V. Dargar, D. A. Yuhas, P. A. Cruickshank, R. Chaguturu, *Eur. J. Biochem.* **2000**, *267*, 394.
[42] J. G. Dann, G. Ostler, R. A. Bjur, R. W. King, P. Scudder, P. C. Turner, G. C. Roberts, A. S. Burgen, *Biochem. J.* **1976**, *157*, 559–571.
[43] K. Iwai, M. Kohashi, H. Oe, *Agric. Biol. Chem.* **1981**, *45*, 113–120.
[44] R. Arrebola, A. Olmo, P. Reche, E. P. Garvey, D. V. Santi, L. M. Ruiz-Perez, D. Gonzalez-Pacanowska, *J. Biol. Chem.* **1994**, *269*, 10590–10596.

Received: June 26, 2007
Revised: October 19, 2007
Published online on December 5, 2007

