Multiple interests in structural models of DARC transmembrane protein.

Dorota Smolarek, Olivier Bertrand, Marcin Czerwinski, Yves Colin, Catherine Etchebest, Alexandre de Brevern

Accepted for publication in *Transfusion Clinique & Biologique* (2010)

Multiple interests in structural models of DARC transmembrane protein.

Multiples intérêts des modèles structuraux de la protéine transmembranaire DARC.

Dorota Smolarek[1,2], Olivier Bertrand[1], Marcin Czerwinski[2], Yves Colin[1], Catherine Etchebest[1] and Alexandre G. de Brevern[1,§]

[1] INSERM UMR-S 665, Université Paris Diderot – Paris 7, Institut National de la Transfusion Sanguine (INTS), 6, rue Alexandre Cabanel, 75739 Paris cedex 15, France.

[2] Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wroclaw, Poland.

Short title: DARC protein.

[§]Corresponding author: Alexandre G. de Brevern, INSERM UMR-S 665, DSIMB, Université Paris Diderot – Paris 7, Institut National de la Transfusion Sanguine (INTS), 6, rue Alexandre Cabanel, 75739 Paris cedex 15, France.

Abstract

DARC (Duffy Antigen Receptor for Chemokines) is unusual transmembrane chemokine receptor which (i) binds the two main chemokine families and (ii) does not transduct any signal as it lacks the DRY consensus sequence. It is considered as silent chemokine receptor, a tank useful for chemiotactism. DARC had been particularly studied as a major actor of malaria infection by *Plasmodium vivax*. It is also implicated in multiple chemokine inflammation, inflammatory diseases, in cancer and might play a role in HIV infection and AIDS. In this review, we focus on the interest to build structural model of DARC to understand more precisely its abilities to bind its physiological ligand CXCL8 and its malaria ligand. We also present innovative development on VHHs able to bind DARC protein. We underline difficulties and limitations of such bioinformatics approaches and highlight the crucial importance of biological data to conduct these kinds of researches.

Keywords: transmembrane protein, comparative modelling, bioinformatics, protein docking, flexible docking, camelid VHHs, malaria, cancer.

Résumé

Le DARC (Duffy Antigen Receptor for Chemokines) est un récepteur aux chimiokines inhabituel qui (i) lie les deux grandes familles de chimiokines et (ii) du fait de l'absence du motif DRY ne transduit pas de signal. Récepteur silencieux, il est un réservoir utile pour le chimiotactisme. DARC a été particulièrement étudié comme un acteur majeur de l'infection par *Plasmodium vivax*. Il est également impliqué dans des maladies inflammatoires, cancers et pourrait jouer un rôle dans l'infection par le HIV. Nous présentons l'intérêt de construire un modèle structural de DARC, pour comprendre plus précisément sa capacité à lier son ligand physiologique CXCL8 et son ligand paludique. Nous présentons des développements innovants portant sur des VHHs capables de lier le DARC. Nous soulignons aussi les difficultés et les limites des approches bioinformatiques et mettons en évidence l'importance cruciale de données biologiques pour mener à bien ce type de recherches.

Mots-clés: protéine transmembranaire, modélisation comparative, bioinformatique, assemblage protéique, assemblage flexible, VHHs de camélidés, paludisme, cancer.

DARC

The history of human knowledge on DARC (Duffy Antigen Receptor for Chemokines) begins in 1950 with the discovery of a new blood groups system (the Duffy blood group system) named from the person who developed the first antibody against the so called Fya antigen [1]. A second antithetic antigen Fyb [2] was shortly after discovered. In 1955, it was shown that antigens of Duffy blood group system were missing in red blood cells (named Fy(a-b-)) from a large proportion of West African ascent population (RBC-WAAP) [3]. It was observed thereafter that these cells were resistant to invasion by *Plasmodium vivax* (see below). Other important steps were cloning of cDNA coding for the protein carrying the Fy antigens, the Duffy glycoprotein, the recognition that Duffy glycoprotein was a transmembrane receptor for chemokines leading to changing its name to DARC.

In this short review, we will briefly overview knowledge on the immunogenic properties of DARC, relations of DARC with malaria, of DARC with chemokines inflammation and inflammatory diseases. We will quote present research which deals with the multiple roles of this somewhat enigmatic protein that, besides malaria and inflammation, is implicated in cancer and might play a role in HIV infection and AIDS. Then, we show the interests in the design of structural models for DARC analysis. We will present (i) how to build proper structural models of DARC [4], (ii) how to elucidate pertinent interactions with its ligands [5] and (iii) what might be the role of structural modelling in elaboration of new tools for DARC studies [6].

Duffy antigens.

They have been defined by studying reactivity of patients immunized through transfusion or pregnancy. Fya/Fyb allotypic variants exist and correspond to a SNP in

exon 2 encoding a Gly42Asp substitution in the extracellular N-terminal domain of the Duffy glycoprotein [7, 8]). Two other antigens have been identified: (i) Fy3, which involves residues from the 3rd extracellular loop [9], is probably a conformational reader and (ii) Fy6, which was discovered after immunisation of mice with human red cells or engineered eukaryotic cells expressing DARC, is a linear epitope contained in the first extracellular domain. Fy6 is present both in Fya or Fyb allotype, and, only Fy(a-b-) cells do not react with anti Fy6.

The mechanism of selective extinction of expression of Duffy related antigens on WAAP red cells have been elucidated. The Duffy negative phenotype of WAAP red cells (noted Fy(a-b-)) is due to homozygosity for a promoter polymorphism (-46C) in which the binding site for the transcription factor (GATA-1), required for DARC to be expressed on the cell surface of erythrocytes [10], is disrupted. This mutation is present in a Fyb genetic background. Importantly in Fy(a-b-) WAAP, DARC is normally expressed on cells in which DARC expression was already demonstrated for example, endothelial cells of post capillary veinules, epithelial cells of collecting ducts of the kidney, cerebellar Purkinje cells [11, 12]. Another promoter is likely operative in these tissues.

DARC and Plasmodium vivax. DARC was characterized as an erythrocyte receptor for malaria parasite through *in vitro* studies and also *in vivo* experiments performed on American volunteering detainees [13, 14]. The hypothesis that DARC might be a receptor for *P. vivax* raised after it was noted that WAAP might be resistant to infection by *P. vivax* purportedly performed to treat neuro syphilis. All this does support the widely accepted hypothesis that *P. vivax* was the driving force for fixing the mutation silencing red cell expression of DARC. In this regards, it is

interesting to note that in Papua New Guinea, where *P. vivax* malaria is also endemic, heterozygous individuals for the same GATA-1 site mutation have been found [15] but on a Fya background. It is tempting to speculate that the same FY GATA-1 mutation in Africans and Melanesians occurred independently in these two populations as a result of the same selection pressure.

Plasmodium vivax Duffy binding protein (PvDBP) is a merozoite microneme ligand vital for blood-stage infection, which makes it an important candidate vaccine for antibody-mediated immunity against *vivax* malaria [16, 17]. Naturally acquired antibodies to DBP seem to confer protection from blood-stage *P. vivax* infection, supporting the development of a vaccine against *P. vivax* malaria [18]. However other studies also pinpointed that produced human antibodies might have low efficiencies underlining the difficulty of vaccine design [19]. Hence, alternative approaches to interfere with *P. vivax* merozoite with DARC on red cells are demanded. Consequently, analysis of interaction mechanisms between DARC and DBP is important; analysis of DBP variants and DARC genotypes gives also insights to the sequence – function relationship [20].

Very recently, studies have shown that in Madagascar, *P. vivax* can invade Fy(a-b-) erythrocytes leading to disease [21]. Further studies are necessary to identify the genetic peculiarities of the parasite strain the receptors that enable this DARC-independent *P. vivax* invasion of human erythrocytes.

Beyond DARC and PvDBP it is worth to notice the existence of Duffy-Binding like (DBL) domains implicated in other types of malaria. Domains related to PvDBP are found in *Plasmodium falciparum*. DBL domains are conserved regions of erythrocyte membrane protein 1 (PfEMP1) family. VAR2CSA Duffy binding-like (DBL) domains, which bind chondroitin sulphate A in placenta, are interesting

candidates for the development of a vaccine against pregnancy-associated malaria [22]. Indeed, in spite of the extreme polymorphism of PfEMP1 DBLα domains, specific antibodies reducing risk of malaria in areas with high transmission rates were acquired [23]. DBLs family fold is supposed to be conserved. Consequently, the family is intensively studied to elucidate binding mechanisms [24-27].

DARC and chemokines. DARC is a transmembrane receptor for a variety of chemokines of both CXC and CC classes, including angiogenic (ELR^+) CXC chemokines, but not angiostatic (ELR^-) CXC chemokines [28, 29]. DARC sequence is quite different from other chemokines receptors [30, 31]. It is a silent chemokine receptors (or interceptors) [29]. Besides, a clear distinction should be made between DARC expressed on red cells and DARC expressed in other tissues. Importantly, DARC is lacking the DRY consensus sequence that is necessary to activate a protein G dependant activation cascade after activation by ligand binding [8, 32-34]. DARC on red cells does not internalize. DARC might play the role of a buffer or a scavenger for chemokines and could reduce their concentration in blood stream [35]. By contrast DARC on endothelial cells behaves differently. It supports transcytosis of chemokines from luminal to extravascular space and favours leucocyte migration and development of inflammatory reactions [36]. A similar mechanism might operate in *in vitro* model of rheumatoid arthritis in which overexpression of DARC does favour inflammatory reaction [37]. DARC could play a role in inflammatory diseases of the kidney [38].

Heterodimerization of DARC with CCR5 might impair activation of intracellular signalling dependant on chemokine binding to CCR5 [39]. This observation definitely adds a level of complexity and makes the role of DARC difficult to understand.

Evidence strongly supports that DARC might play role(s) in the delicate interplay of chemokines and chemokines receptors (chemokines network). DARC is definitely considered to be important in a range of physiological and pathological conditions.

DARC and HIV. HIV binding was early reported [40, 41]This binding, which is inhibitable by CXCL8 8 and anti FY6 antibody [41], might help the viral infection of CD4+ sensitive cells. Later results confirmed that red cell binding to HIV might favour HIV trans-infection of susceptible cells. However no difference was observed when comparing Fy(a+b+) and Fy(a-b-) cells [42]. Epidemiological studies did suggest that the Duffy-null state is associated with a survival advantage in HIV-infected persons of African ancestry [41, 43], but this proposition is also still highly debated [44-46].

DARC and cancer. Recently, a relation between DARC and various cancers was established, making DARC a hot topic. Hence, it seems that DARC and murine CXC Chemokine Receptor-2 Receptors have opposite role in murine melanoma tumor growth [47, 48]. Epidemiological studies showed that African Americans are suffering from prostate cancer earlier than Caucasians. Moreover, the course of the disease appears more aggressive for African American population. These observations suggested that DARC might play a role in cancer. These observations were backed up by crossing DARC null engineered mice with TRAMP mice which developed spontaneous prostate cancer [28]. The scavenger role of DARC for chemokines to excess in red cells could participate in reducing angiogenesis, and consequently the progression of prostate cancer, by clearing angiogenic chemokines from the tumour

microenvironment [28, 49]. Hypothesis has been discussed on the substantially higher levels of prostate cancer and associated mortality rates in men of African origin compared with Caucasian men. On the other hand, expression of DARC in breast cancer tumour cells does lower metastasis and aggressiveness of the tumour [50]. Again, interaction of DARC with chemokines is hypothesized to explain a putative role of DARC in cancer progression noticeably angiogenic chemokines that are obviously implicated in tumour neovascularisation.

The different aspects of the research

All these data clearly suggests that DARC is a particularly interesting and important protein. Various DARC mutants have been designed and expressed. Affinities with DARC natural ligand, CXCL8, and different antibodies [33, 51-58] confirmed predictions made about DARC topology [59]. This transmembrane protein as *bona fide* GPCR has 7 transmembrane segments with four extracellular loops (named Extra Cellular Domains or *ECDs*) and four intracellular loops (named Intra Cellular Domains or *ICDs*). The first *ECD* (*ECD1*) is a long Nterminal segment, while last *ICD* (*ICD4*) is a short cytosolic Cterminal.

Structural information can help a lot to understand its function and implications in diseases [60]. However, few transmembrane protein structures (~ 650 structures, 1% of the Protein Databank [61-63]) are nowadays available [64, 65]. Due to the membrane environment [66-69] that stabilises the 3D fold, it is particularly difficult to extract, to purify to crystallise and finally to solve transmembrane proteins structures by X-ray crystallography. Hence, structural modeling is an obligatory but difficult step. With a low number of available 3D structures, automatic homology modeling cannot be applied to GPCR or GPCR-like molecules even though it was attempted: at

best position of transmembrane membrane domains may be predicted but no solid information on structure of loops may be gained. Thus, pertinent structural models of transmembrane proteins must be a human supervised process using classical approaches coupled with various secondary structure predictions, refinement of sequence alignment, and if possible considering helix – helix and helix – lipid interactions. Importantly, experimental data are necessary for building relevant models [70].

In the next section, we will detail a concrete application and explain how experimental data were crucial for building pertinent structural models [4]. These models were particularly helpful to explain DARC interaction with its ligands. We will present the principle of docking methods applied to this kind of protein and the specific difficulties met when dealing with DARC [5]. Finally, we will discuss the interest of proposing structural models of camelids VHHs shown to bind a specific DARC epitope [6].

Building structural models

Classically, a structural model can be elaborated through different strategies from homology/comparative modelling, threading *ab initio* or de novo approaches, depending on the sequence identity and the availability of structural homologous. Figure I shows a rough description of the sequence identity needed for each of these approaches. If the sequence identity is high, homology modelling could be used. In the twilight zone, when PSIBLAST is unable to detect any interesting sequences, threading approach could help to find distant homology by evaluating sequence-template structure fold compatibility. In case of failure, *ab initio* modelling becomes the only possibility. *Ab initi*o methods do not require any 3D template, but physico-

chemical and/or statistical potentials. The main problem consists in finding the native protein structure, *i.e.* the 3D structure with a minimum free energy. Some *ab initio* approaches try to mimic protein folding to find this minimum. These approaches are essentially successful for peptides and very small proteins. Otherwise, they are not performed due to search complexity. *De novo* approaches combined these different methods starting from domain detection, looking for fragments with similar sequences in PDB. A large set of fragments is generally considered. The next problem consists in appropriately combining these fragments. The combinatory becomes rapidly prohibitive as the size of the protein increases. Consequently, sophisticated algorithms aiming at solving this combinatorial problem are required. These methods give good results but need an important computer power. These different approaches are available on meta-servers that mainly combine multiple results from different softwares and/or servers [71, 72]. It has to be noticed that all these approaches were developed for globular proteins.

For the building of DARC of structural models, we started with comparative modelling. The procedure first consists in searching for homologous sequences in PDB. However, classical tools did not provide any convincing answer. The sequence identity between DARC and rhodopsin (PDB code: 1F88 [73]), the paradigm for 7-TM fold, was very low, *i.e.*, only 12% in the range of random alignment). Consequently, sequence alignment was not meaningful and no clear homology could be inferred. Moreover, DARC family members were too few to detect important conserved residues from divergent ones with the sole alignment. Consequently, the absence of real sequence divergence makes the alignment not truly informative. Actually, pertinent structural models could not be properly built with the sole sequence information of DARC. However, for DARC, important biological data were

available. Indeed, more than 40 different punctual DARC mutants were experimentally tested for affinity with natural ligands or some antibodies [9, 53, 74], they underline the potential accessibility of some residues. With this information, it was possible to guide the building of structural models. Figure II presents the main principle of the structural model building.

DARC was divided into three main domains: (i) *ECD1*, (ii) one central domain encompassing the transmembrane helices and connection loops, and (iii) *ICD4*. Putative helical transmembrane regions were located with dedicated prediction softwares (DAS [75], TopPred 2 [76, 77], HMMTOP [78, 79], TMHMM 1.0 and 2.0 [80], PHDtm [81, 82], TMpred [83], SOSUI [84-86], SPLIT [87, 88], Pred-TMR 1.0 and 2.0 [89, 90], TMAP [91, 92], TSEG [93], TM-FINDER [94], UMDHMM^{TMHP} [95], MEMSAT [96, 97], PRODIV-TMHMM [98] and MemBrain [99]). All these methods claim to be efficient with significant prediction rates when tested on benchmark datasets. Only the first helix is predicted with a large consensus. The fourth helix was particularly difficult to delimit. Other helices could also diverge by an impressive number of residues, *i.e.*, 15 residues. The most recent prediction tools, *e.g.*, MINNOU [100] were not necessarily the most efficient ones. Indeed, in some cases, PSI-PRED [101] mainly trained on globular protein, could give better results than dedicated approach, as seen in [102].

In a second step, using a rough consensus definition of transmembrane regions, predicted helices were aligned with assigned transmembrane helical (DSSP software [103])segments of rhodopsin (PDB code 1F88 [73]). Strong efforts were dedicated to the prediction of *ECD1* and *ICD4* [104, 105]. Hundred models were generated using Modeller software [106-108]. Each model was then refined: (i) the side chains were repositioned using one of the most efficient approach, *i.e.*, SCWRL [109]. We

performed simulated annealing for exploring connection loops conformations using GROMACS software [110, 111]. The residue accessibility was computed with Naccess software [112] and the results were compared to experimental data available. We focused on residues involved in antibody binding that are supposed to be accessible. The alignment was then modified accordingly. Twelve generations of alternative alignments were tested for finally obtaining two structural models that diverge by the topology of *ECD1*. (see Figure III [113]). In these two models, the accessibility values of important residues were large enough to allow binding.

Since, novel approaches have been developed and adapted to transmembrane proteins. We revisited our models at the light of the most efficient new tools and compared with our results. For this purpose, we tested LOMETS (LOcal MEta-Threading-Server [114]), a webserver that uses 8 different methods and ranks the results. Table I summarizes the different results obtained for MUSTER [115], SAM [116], PROSPECT2 [117], SP3 [118], PPA-I, HHsearch [119], SPARKS2 [120], and FUGUE [121]. The three first methods provide structural models with a medium confidence rate while the models constructed with the last methods are associated with a low confidence index. Only half of the proteins used as template are transmembrane proteins, half are globular proteins with often beta-sheets. Figure IV describes the main results obtained from LOMETS [114] and from PHYRE [122].

Figures IVa, IVd and IVg show the three templates found by LOMETS for the medium scored structural models. The first one is based on the famous human A2A adenosine receptor (PDB code 3EML [123]), while the two others are globular proteins, namely the COPI gamma-subunit (PDB code 1PDZ [124]) and cell vibrio mixtus mannosidase 5A (PDB code 1UUQ [125]). Figures IVb and IVc show the final structural model obtained from A2A adenosine receptor template. The fold

corresponds to a seven-segment transmembrane protein with *ECD1* in an extended conformation. In both models, the first helix is well predicted, but many differences are observed in the positions of the following helices. Nevertheless, these models are good starting points to build pertinent structural models. For the structural model based on COPI gamma-subunit (see Figures IVe and IVf), the difficulty to obtain transmembrane protein model is clearly illustrated. Finally, no structural model based on cell vibrio mixtus mannosidase 5A was obtained, *i.e.*, the alignment being too poor. PHYRE, a powerful threading approach, proposed the well-known bovine rhodopsin structure as a template (PDB code 1U19 [126]) associated with 60 % confidence rate. However, in this structural model presented on Figure IVh, the TM helices were too long or too short helices, compared to what is classically expected for transmembrane protein. The 7-segment transmembrane topology was finally lost.

These revisited models obtained with up-to-date methods show the importance of considering biological data to produce pertinent structural models. It also reinforces the validity of the structural models we constructed some years ago.

Structural properties of DARC

To explore the flexibility of the *ECD* loops, we performed simulated annealing simulations [127]. Interestingly, the procedure highlights the importance of residue D263 which was never really accessible in any structural models; this residue constrains the local fold by creating a bridge with *ECD3*. Analysis of simulations with Protein Blocks [104, 128] showed that that some regions in *ECD1* tent to be more helical and other ones to be more extended. These results correlated well with the predictions done [105, 129, 130].

The two extreme positions of *ECD1* (see Figure III) could also reflect also the domain motions [60]. We performed normal mode analysis (NMA) using different webservices. This methodology has been recently re-popularized and successfully applied for examining dynamics of large systems and also for transmembrane proteins [reference]. Among the different webservers, WEBnm@ server [131] provides additional analysis dedicated to transmembrane protein. For this review, we also tested Nomad-Ref [132] and ElNemo [133, 134]. In NMA, the lowest frequencies modes are associated with the largest amplitude of motions. a large domain motion of *ECD1* that gets closer to the other ECDs was observed with the different NMA tools. A similar motion was indeed observed with Nomad-Ref [132] or ElNemo [133, 134].The main difference with Webnm@ lies in the ranking of the motions with respect to the frequency that were slightly different depending on the tool. The three regions detected with simulated annealing (a first structured zone, a transition region and a last structured zone) are clearly involved in the motion. The median region plays a role of hinge between the two extremities of *ECD1*. Figure V illustrates the motions associated with the two of the lowest modes given by ElNemo [133, 134]. On the left side is indicated the hinge region. The movements are schematized in the two following Figures. In both, the flexible hinge region moves with respect to the two structured regions.

Electrostatic potentials of DARC model and its natural ligand CXCL8 were calculated using the finite-difference Poisson-Boltzmann (FDPB) method [135]. Two distinct zones can be observed in potential interaction zones with CXCL8. The first one is highly negative and encompass the residues implicated in epitope Fy6. The second one is highly positive. CXCL8 also shows two regions with opposite features (one positive and the remainder negative, encompassing the loop 40s known to

interact with chemokine receptor). Our results highlight regions with significant electrostatic properties in agreement with experimental studies that underline the importance of electrostatics in the binding of chemokines [136, 137].

A following question is the potential use of such structural models for understanding the binding of DARC and its ligands. Docking methods are particularly appropriate to locate binding sites on both partners and their relative orientation, even though the use of models increases the risk of obtaining irrelevant structures of complex. Docking procedure roughly consists in moving the smallest partner (designed as "the ligand") on the surface of the largest molecule ("the receptor"). For each position, a score (or energy) is computed. The position with the optimal score is finally selected. A source of errors comes from the inadequacy of scoring to represent binding energies. In addition, most docking at least in the first steps, consider the partners as rigid. This limitation is only alleviated in a final refining step when a subset of solutions has been already established. In the case of DARC, we previously underlined that *ECD1* is highly flexible. Clearly, this property must be accounted. Consequently, we performed the study in two steps and we designed a docking approach that combines rigid and flexible docking. In a first step, DARC structural model is cut into *ECD1* and the rest of the protein (see Figure VIa) to find correct positioning of CXCL8 on DARC without *ECD1*, and a flexible docking only with ECD1 and CXCL8 (see Figure VI). On one side, a flexible docking of *ECD1* is done with structure of CXCL8 (see Figure VIb) thanks to ICM [138, 139] software. It is very difficult and complex approach with a very high computing consuming time. On the other side, a rigid docking is done with the transmembrane domain of DARC (see Figure VIc) thanks to ClusPro [140, 141] webserver. Each experiment give numerous possibilities, the docking conserved where selected on energetic properties of the

complex and also using biological data. Figure VII shows examples of results of rigid docking (see Figure VIIa and VIIb) and of final combination of rigid and flexible docking (see Figure VIIc and VIId). These results were quite encouraging. Finally, both results will be combined to perform a deeper search [60]. Further optimization needs to be done, but conserved results are in accordance with expected residues in contacts.

Modelling of camelid VHHs

DARC is implicated in numerous human diseases. Dedicated tools are demanded for analyzing DARC role and guiding therapeutic strategies. In this field, antibodies and their recombinant derivatives are of great use. The heavy chain-only antibodies found in camelids are composed of heavy chains and lack all light chains [142]. VHHs (or nanobodies), which correspond to the domain in the heavy chain-only antibody, can be derived. In this domain is located the antigen recognition region. VHHs are easily cloned from lymphocytes from naive or immunized camelids; they can be expressed in *E. coli* with a good yield and have an excellent solubility [143]. Moreover, they have proved to be efficient as therapeutic and diagnostic agents [144, 145].

A dromedary has been immunized with *ECD1* of DARC expressed in *E. coli*. As presented earlier, *ECD1* carries several sites important to DARC functions and properties (binding to chemokines and PvDBP, Fya/Fyb allotypes, the Fy6 epitope). A VHH library from dromedary's lymphocytes was built and screened using also E. coli expressed proteins for DARC specific VHHs. Several clones were obtained, especially one named CA52. CA52 is able to recognize the glycosylated protein present on human cells, even if the constructs used for immunization and screening

was non glycosylated proteins. The linear epitope recognized by CA52 was identified and overlaps the well known Fy6 antigen. CA52 interferes with the CXCL8 binding to DARC and *P. vivax* infection of red blood cells [6].

To more deeply analyze CA52 and to get a glimpse into its putative interactions with DARC, we constructed a structural model of CA52 VHH. A classical comparative modelling approach was carried out as available VHH structures sharing a good sequence identity with CA52 are available. From a structural point of view, VHHs adopt a well-characterized topology composed of four very constant regions and three hypervariable regions (*CDR1* to *CDR3*). These last ones correspond to the binding part of VHHs that recognize the epitope (here *ECD1*). Constant region fold corresponds to a series of β-sheets that is found in all VHHs and gives the topology of the protein. Figures VIIIa and VIIIb shows an example of a VHH [146], in green the constant regions, and in yellow, orange and red, *CDR1* to *CDR3*. Figures VIIIc highlight the protein interface which binds the epitope.

Sequences related to CA52 were searched with PSI-BLAST software [147] applied on Protein DataBank [61]. Using default parameters of PSI-BLAST, one VHH (PDB code 1XFP [148]) was selected with a very good sequence identity (75%). However, all the structural models obtained presented a topology inversion between two loops. A careful analysis of PSI-BLAST results showed that *CDR3* regions were considered as non-informative (*i.e.*, coiled-coil as detected by SEG [149]) although this *CDR* is the most important one for the binding. When SEG was disabled, PSI-BLAST search gave VHHs structural templates better matching for CDR3 (both in length and sequence identity). A first structure (PDB code 1OP9 [146]) was selected (a sequence identity of 75% and a good alignment with *CDR3*). Figure VIIId underlines the huge impact of the fine analysis of VHHs. *CDR3* of VHH

1XFP is shown in yellow and exhibits a classical β-sheet, while in 1OP9 template structure, *CDR3* of VHH (shown in red) has helical tendencies (mainly a succession of β-turns). Moreover, these regions are totally differently positioned. This example illustrates the importance of a precise analysis of templates and of a good knowledge of bioinformatic tools.

A very important point for guiding the structural modelling was the experimental characterization of an extra-disulfide bridge between Cysteines 33 and 107. This disulfide bond is of major biological importance. We considered a second template to 1OP9, namely 1JT0 [150] that possesses a similar extra-disulfide bridge. The protein sequences were aligned with Clustalw2 software [151] and some manual changes were done. The model construction was performed with the Modeller software [107]. Final structural models showed few differences as constant regions strongly constrained the topology and the extra-disulfide bridge constraints strongly *CDRs*. Topology was assessed using ProCheck software [152]. Figure VIIIe shows the final model. *CDR3* is composed of series of β-turns (the disulfide bridge is in blue colour). Figure VIIIf shows the distribution of charges at the surface. *CDR2* and *CDR3* present a strong positive surface (blue colour) as the central part of *CDR1*. Molecular modelling of another VHH sequence that binds the same epitope but with lower affinity gives some hints about important residues. The presence of a positive surface in this region and the presence of a negative surface in the linear target peptide (see *Structural properties of DARC* section) are suggestive for important electrostatic interactions.

Conclusions and Perspectives

Hence, in this short review, we have presented the biological importance of DARC protein and its multiple implications in human diseases. We have underlined the interest to use structural models to better understand this protein. The building of a structural model for a transmembrane protein is a very difficult task. Comparison with up-to-date methods highlights the crucial value of biological data to produce pertinent structural models, our approach remaining the most efficient one. Using these models, we have shown the capabilities of new and complex methodology combining classical rigid docking and novel flexible docking. Additional simulations must be done to confirm these preliminary results, but these last one are already quite encouraging, with a very good agreement with experimental data. Finally, we have opened new perspectives given by the use of structural models of camelid VHHs able to bind DARC. We plan to analyze each sequence of VHHs obtained in a similar way to understand more precisely the most important residues involved in the recognition of *ECD1*.

Acknowledgments

D.S. was supported by scholarship from French Embassy in Warsaw and from the Flemish government. This work was supported by grants from the French Ministry of Research, Université Paris Diderot – Paris 7, French National Institute for Blood Transfusion (INTS), French Institute for Health and Medical Research (INSERM). Project was supported by grant no. N N302 118835 from the Ministry of Science and Higher Education of Poland and in part by a Polonium Partenariat Hubert Curien grant.

Figures

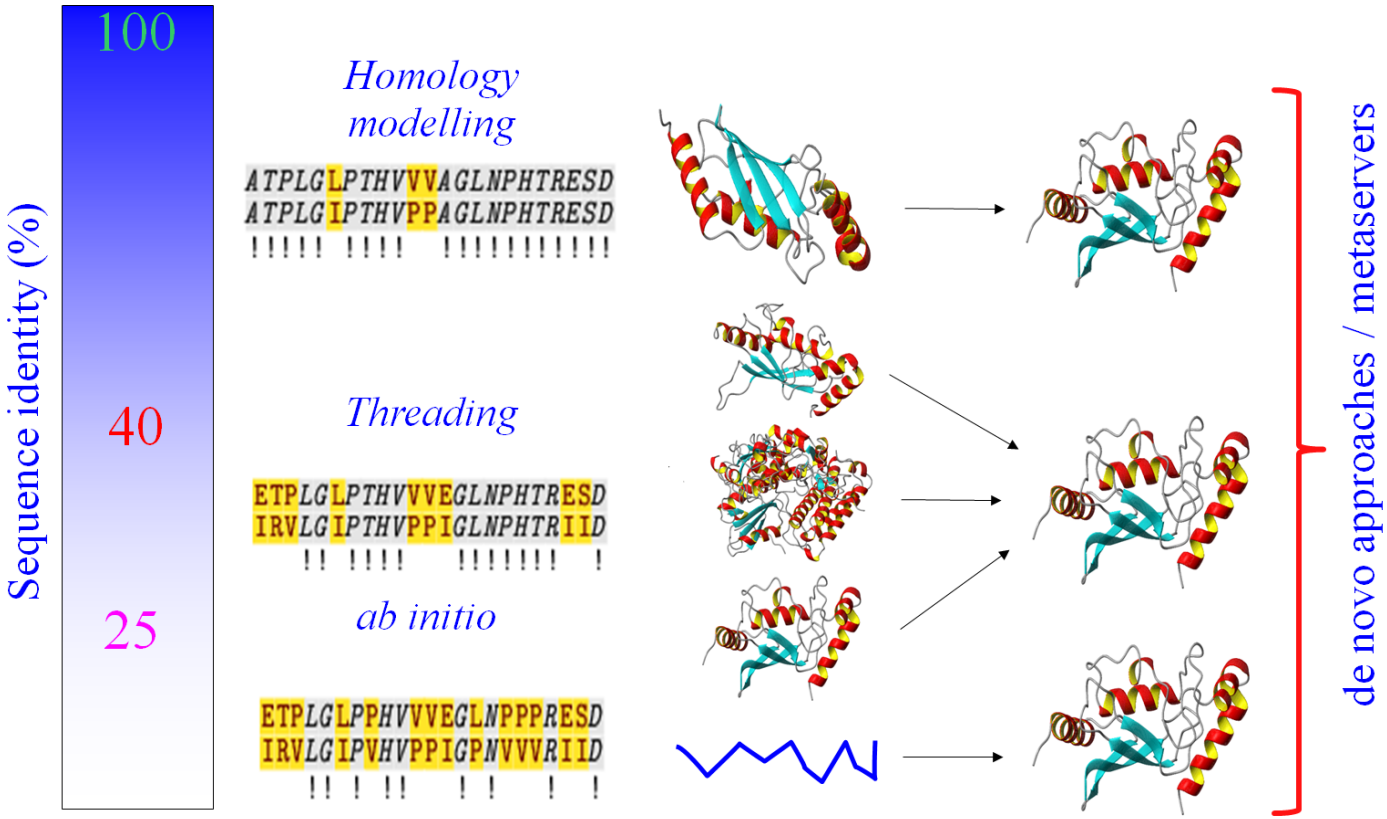

Figure I. *Structural modeling methodology in function of sequence identity.*

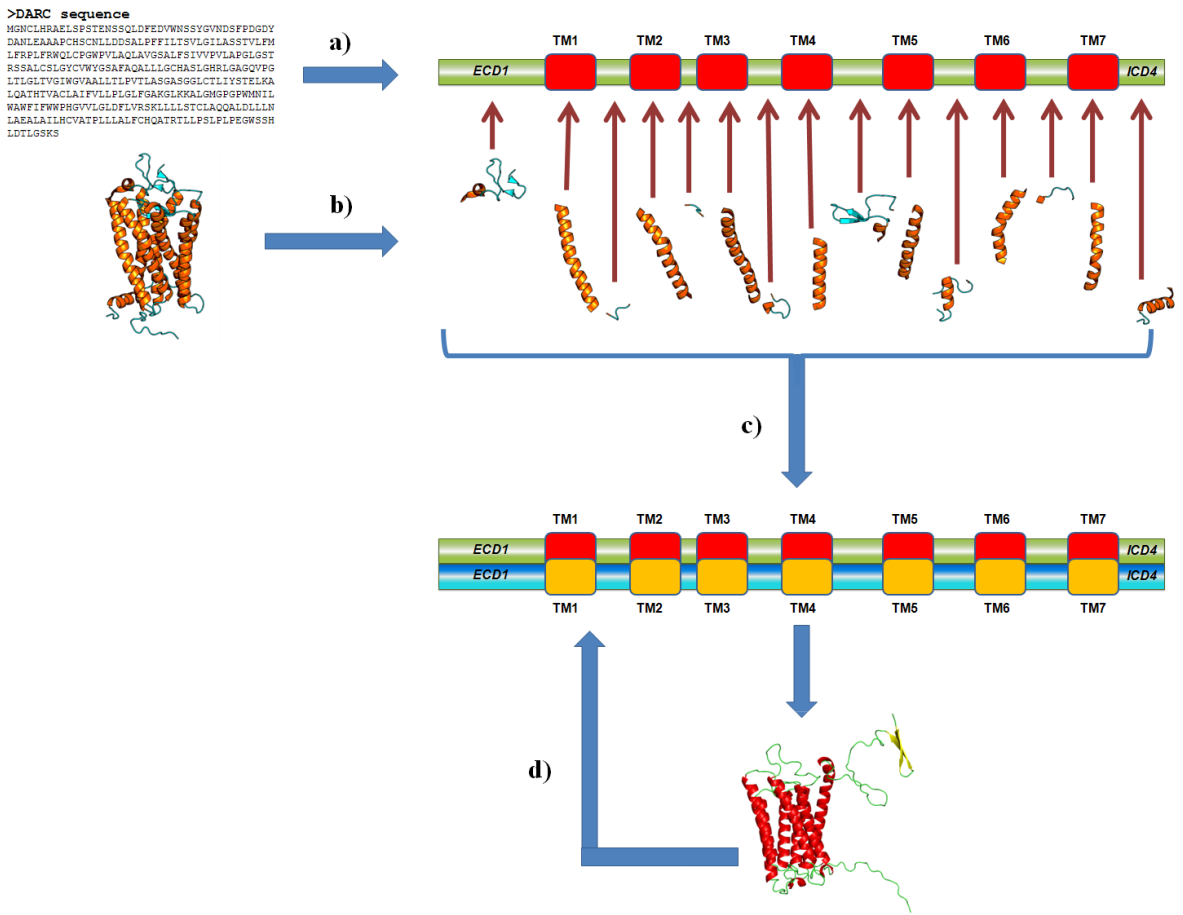

Figure II. *Principle of DARC structural model building*. (a) Prediction of transmembrane segments, the DARC sequence is divided into 7 transmembrane segments and 8 loops. (b) The rhodopsin structure is also cut into transmembrane segments and loops. (c) Each segment is aligned independently with its counterparts. *ECD1* and *ICD4* are treated separately with other approaches (see text). A global alignment is done and used to generate structural models. (d) Structural models are optimized and important residues are manually analyzed. The alignment is then manually modified and new structural models are generated. The process is done until most of the concerned residues are accessible.

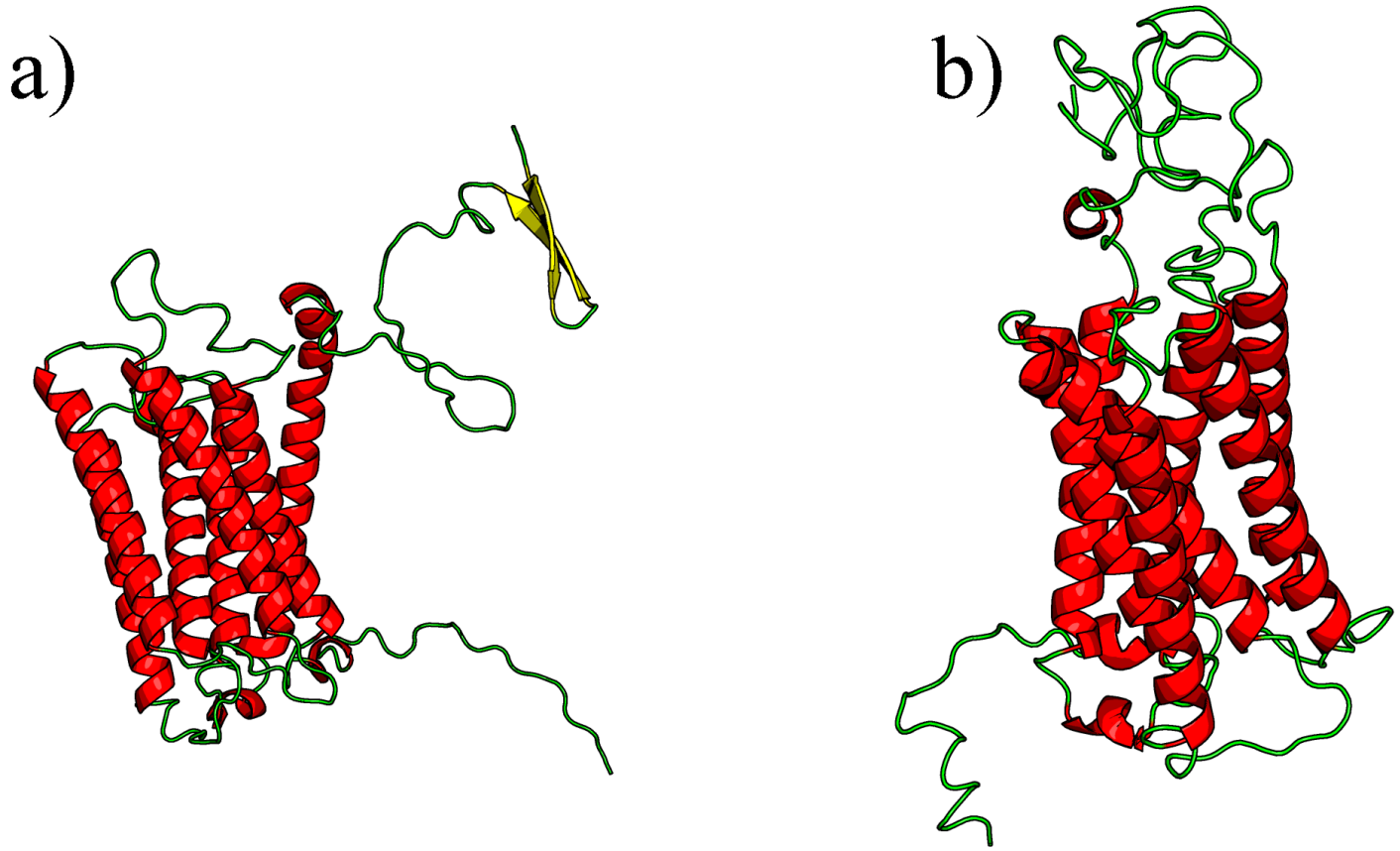

Figure III. *The two selected structural models.* They are shown thanks [113] to PyMol software . (a) open form, (b) closed form.

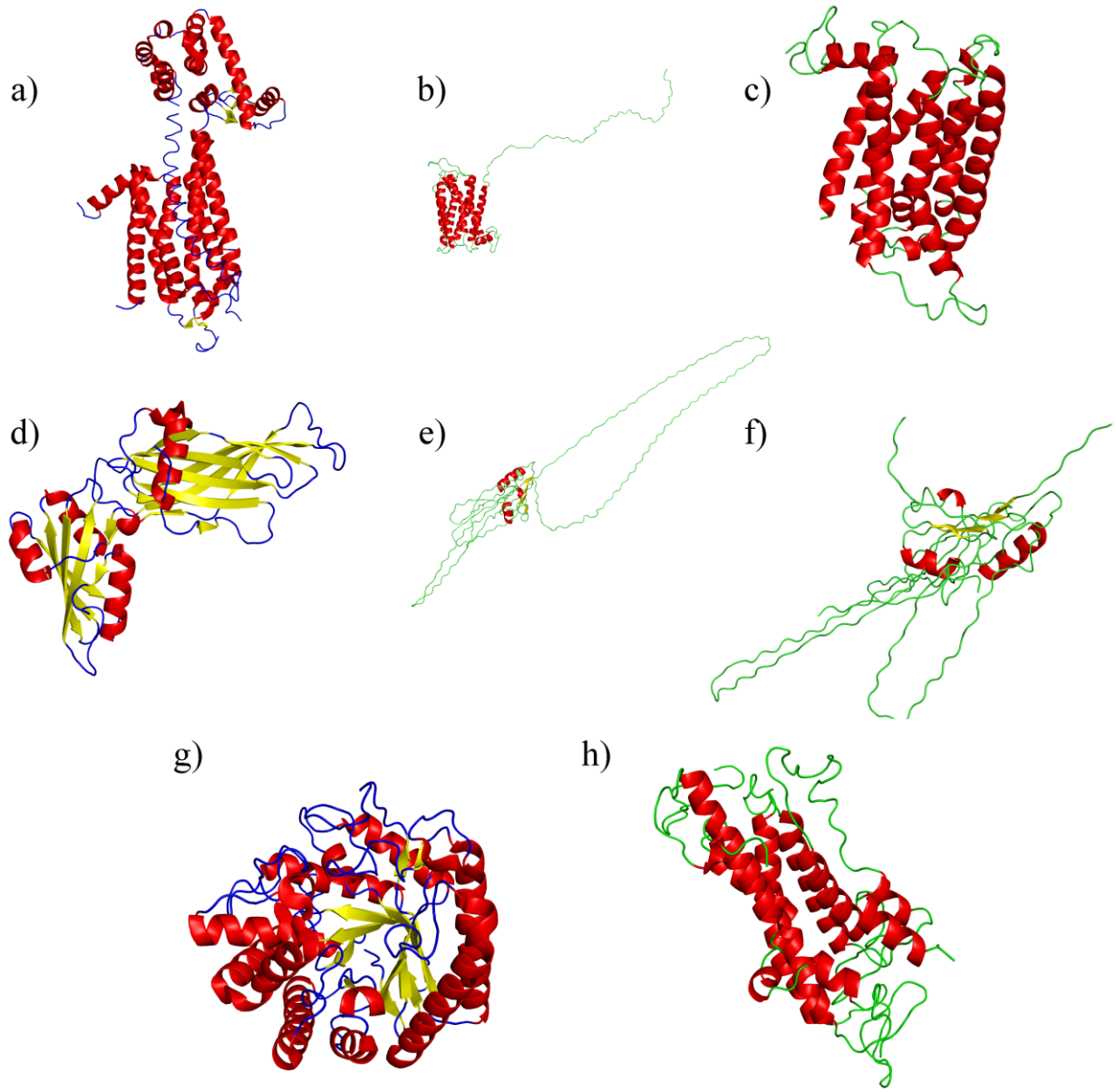

Figure IV. *Results of LOMETS and FUGUE webservers*. (a), (d) and (g): structural templates used by LOMETS [114] for its medium quality structural models. (a): human A2A adenosine receptor (PDB code 3EML [123]), (d) COPI gamma-subunit (PDB code 1PDZ [124]), (g) cell vibrio mixtus mannosidase 5A (PDB code 1UUQ [125]). (b) and (c) two views of the structural model based on human A2A adenosine receptor (PDB code 3EML [123]), (e) and (f) two views of the structural model based on COPI gamma-subunit (PDB code 1PDZ [124]). (h) the structural model based on bovine rhodopsin structure (PDB code 1U19 [126]) obtained by PHYRE [122].

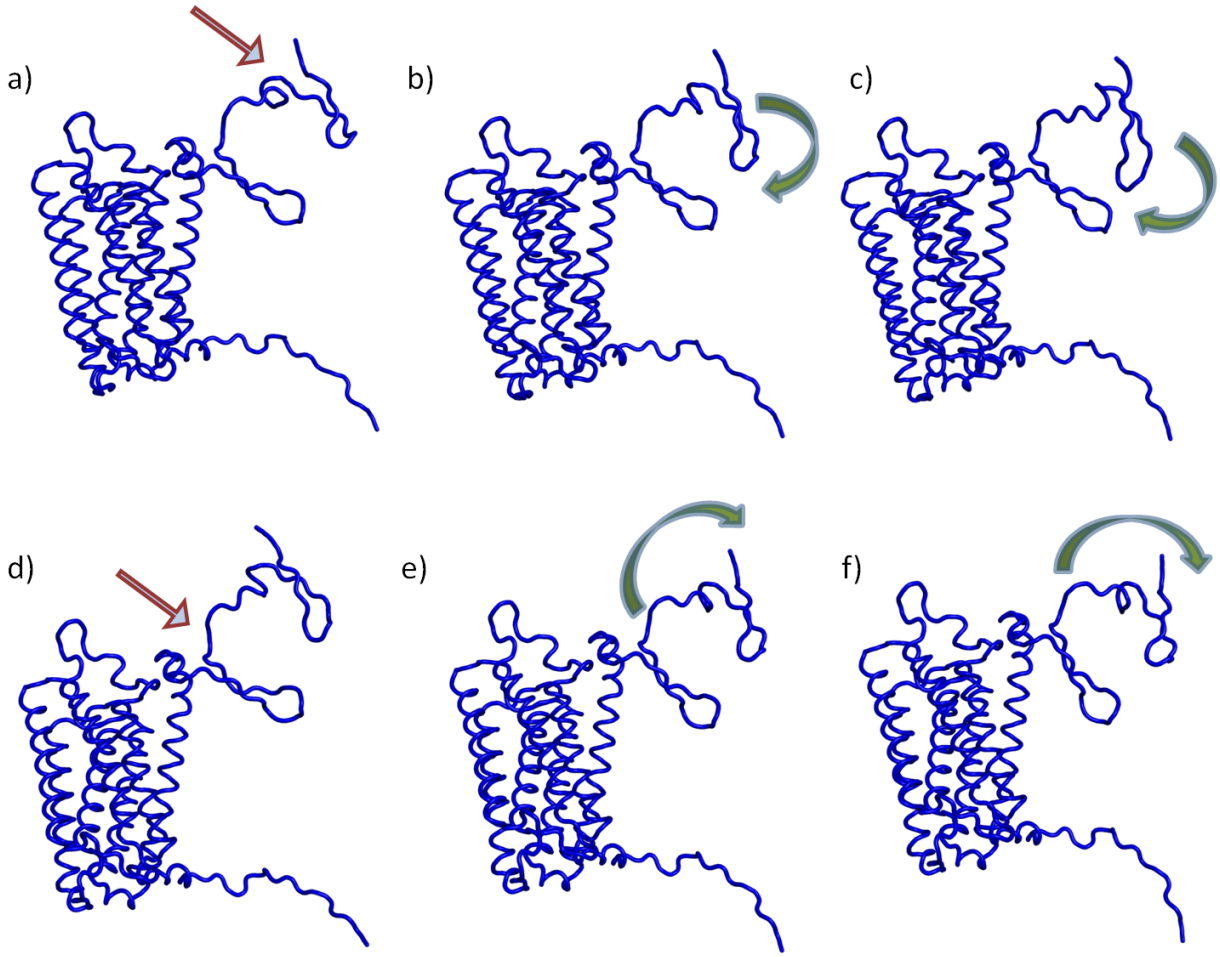

Figure V. *Two low modes computed with ElNemo*. (a) and (c) initial structure, the hinge is pointed by the arrows, (b-c) and (e-f) motions observed with these modes, the arrows indicated the nature of the observed movements.

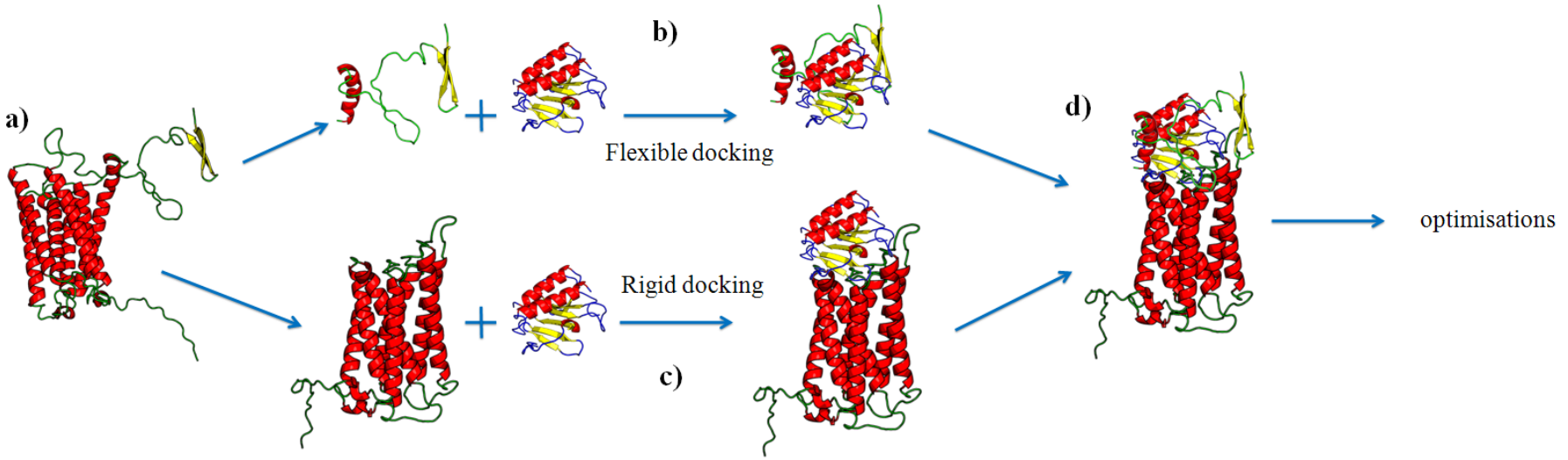

Figure VI. *Principle of our docking approach*. (a) the structural model of DARC is split into two parts (*ECD1* and transmembrane region). (b) flexible docking is performed between *ECD1* and CXCL8. (c) rigid docking of CXCL8 is done with transmembrane domain of DARC. (d) best results of each approaches are combined and then optimized.

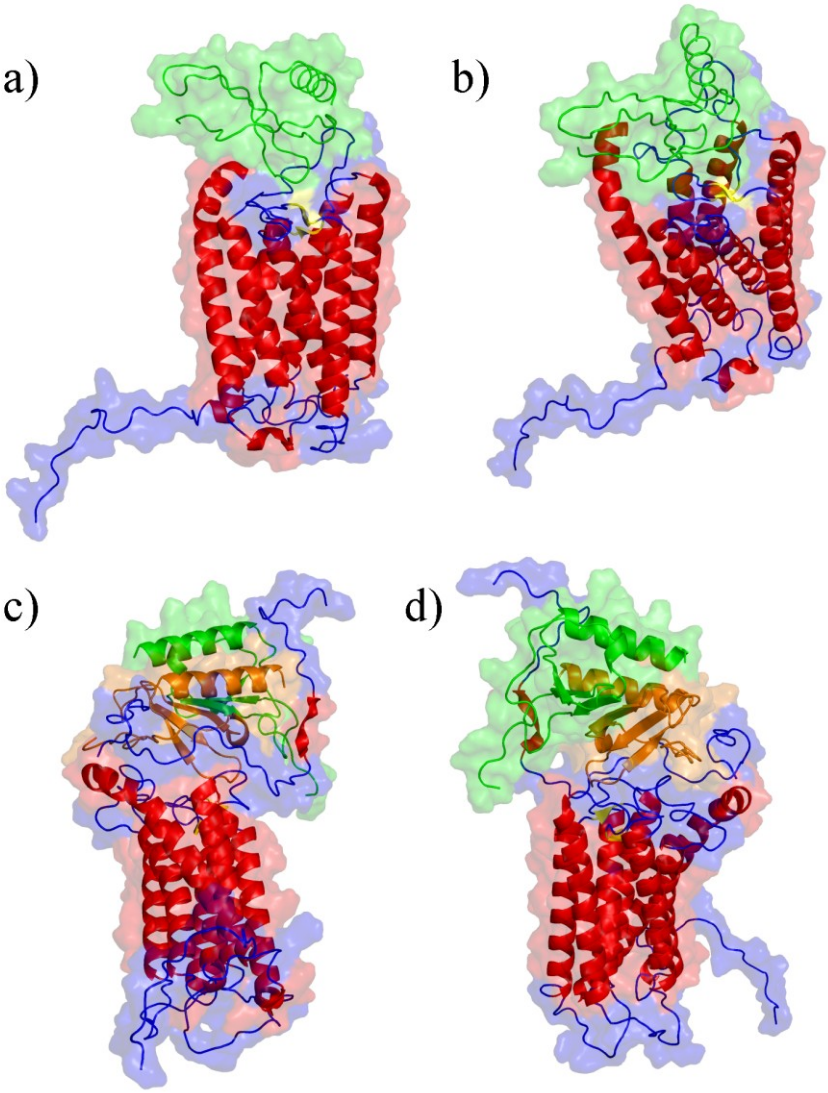

Figure VII. *Two examples of docking results*. (a) and (b) two views of an example of a rigid docking of monomeric CXCL8 (green color) with transmembrane domain of DARC (helices in red, loops in blue and beta-strand in yellow). (c) and (d) two views of an example of a combination of rigid and flexible docking of dimeric CXCL8 (green and orange color) with transmembrane domain of DARC.

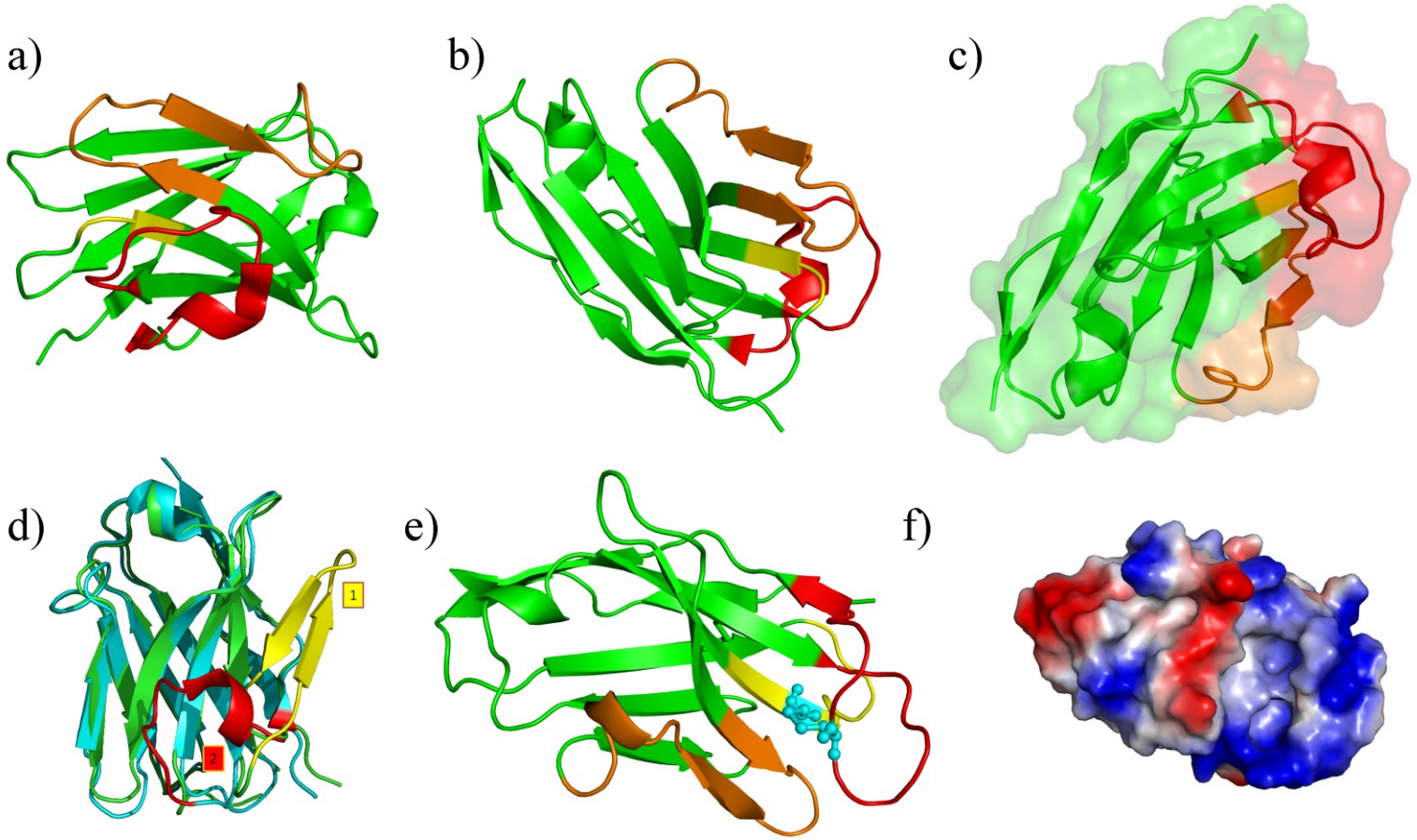

Figure VIII. *Camelid VHHs and structural model*. (a) and (b) two views of an VHH, *CDR1* is in yellow, *CDR2* in orange and *CDR3* is red, the rest is in green. (c) surface representation of the same protein. (d) Comparison between the two potential templates, noted (1) is the original one which corresponds to β-sheet conformation (in yellow), while (2) is the correct one (in red). (e) Selected structural model, the extra-disulfide bridge is in blue. (f) Electrostatics surface.

Tables

Rank	Template	Align_length	Coverage	Zscore	Seq_id	Confidence Score	Program
1	3emlA1	255	0.758	5.444	16.5	Medium	MUSTER
2	1pzdA2	106	0.315	11.656	27.4	Medium	SAM
3	1uuqA	313	0.931	4.622	8.3	Medium	PROSPECT2
4	1l9ha	299	0.889	8.893	14.0	Low	SP3
5	2vt4B	248	0.738	6.260	14.1	Low	MUSTER
6	3emlA1	254	0.755	6.714	15.7	Low	PPA-I
7	2z73_A	269	0.800	4.923	13.4	Low	HHsearch
8	1l9ha	297	0.883	3.021	15.8	Low	SPARKS2
9	hs2f82a	334	0.994	3.070	11.7	Low	FUGUE
10	1wtbA	79	0.235	9.059	20.3	Low	SAM

Table I. *Results of LOMETS webserver.*

References

1. Cutbush M, Mollison PL. The Duffy blood group system. Heredity. 1950;4:383-9.
2. Ikin EW, Mourant AE, Pettenkofer HJ, Blumenthal G. Discovery of the expected haemagglutinin, anti-Fyb. Nature. 1951;168:1077-8.
3. Sanger R, Race RR, Jack J. The Duffy blood groups of New York negroes: the phenotype Fy (a-b-). Br J Haematol. 1955;1:370-4.
4. de Brevern AG, Wong H, Tournamille C, Colin Y, Le Van Kim C, Etchebest C. A structural model of a seven-transmembrane helix receptor: the Duffy antigen/receptor for chemokine (DARC). Biochim Biophys Acta. 2005;1724:288-306.
5. de Brevern AG, Autin L, Colin Y, Bertrand O, Etchebest C. In silico studies on DARC. Infect Disord Drug Targets. 2009;9:289-303.
6. Smolarek D, Hattab C, Hassanzadeh-Ghassabeh G, Cochet S, Gutiérrez C, de Brevern AG, et al. A recombinant dromedary antibody fragment (VHH or nanobody) directed against human Duffy Antigen Receptor for Chemokines. Cellular and Molecular Life Sciences. 2010:in press.
7. Hadley TJ, Peiper SC. From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. Blood. 1997;89:3077-91.
8. Tournamille C, Le Van Kim C, Gane P, Cartron JP, Colin Y. Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. Hum Genet. 1995;95:407-10.
9. Tournamille C, Filipe A, Wasniowska K, Gane P, Lisowska E, Cartron JP, et al. Structure-function analysis of the extracellular domains of the Duffy antigen/receptor for chemokines: characterization of antibody and chemokine binding sites. Br J Haematol. 2003;122:1014-23.
10. Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet. 1995;10:224-8.
11. Peiper SC, Wang ZX, Neote K, Martin AW, Showell HJ, Conklyn MJ, et al. The Duffy antigen/receptor for chemokines (DARC) is expressed in endothelial cells of Duffy negative individuals who lack the erythrocyte receptor. J Exp Med. 1995;181:1311-7.
12. Horuk R, Martin A, Hesselgesser J, Hadley T, Lu ZH, Wang ZX, et al. The Duffy antigen receptor for chemokines: structural analysis and expression in the brain. J Leukoc Biol. 1996;59:29-38.
13. Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK. Erythrocyte receptors for (Plasmodium knowlesi) malaria: Duffy blood group determinants. Science. 1975;189:561-3.
14. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med. 1976;295:302-4.
15. Zimmerman PA, Woolley I, Masinde GL, Miller SM, McNamara DT, Hazlett F, et al. Emergence of FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea. Proc Natl Acad Sci U S A. 1999;96:13973-7.

16. Chootong P, Ntumngia FB, VanBuskirk KM, Xainli J, Cole-Tobian JL, Campbell CO, et al. Mapping epitopes of the Plasmodium vivax Duffy binding protein with naturally acquired inhibitory antibodies. Infect Immun. 2010;78:1089-95.
17. Rowe JA, Opi DH, Williams TN. Blood groups and malaria: fresh insights into pathogenesis and identification of targets for intervention. Curr Opin Hematol. 2009;16:480-7.
18. King CL, Michon P, Shakri AR, Marcotty A, Stanisic D, Zimmerman PA, et al. Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage Plasmodium vivax infection. Proc Natl Acad Sci U S A. 2008;105:8363-8.
19. Souza-Silva FA, da Silva-Nunes M, Sanchez BA, Ceravolo IP, Malafronte RS, Brito CF, et al. Naturally acquired antibodies to Plasmodium vivax Duffy binding protein (DBP) in Brazilian Amazon. Am J Trop Med Hyg. 2010;82:185-93.
20. Storti-Melo LM, de Souza-Neiras WC, Cassiano GC, Joazeiro AC, Fontes CJ, Bonini-Domingos CR, et al. Plasmodium vivax circumsporozoite variants and Duffy blood group genotypes in the Brazilian Amazon region. Trans R Soc Trop Med Hyg. 2009;103:672-8.
21. Menard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, Ratsimbasoa A, et al. Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc Natl Acad Sci U S A.107:5967-71.
22. Avril M, Hathaway MJ, Cartwright MM, Gose SO, Narum DL, Smith JD. Optimizing expression of the pregnancy malaria vaccine candidate, VAR2CSA in Pichia pastoris. Malar J. 2009;8:143.
23. Mayor A, Rovira-Vallbona E, Srivastava A, Sharma SK, Pati SS, Puyol L, et al. Functional and immunological characterization of a Duffy binding-like alpha domain from Plasmodium falciparum erythrocyte membrane protein 1 that mediates rosetting. Infect Immun. 2009;77:3857-63.
24. Higgins MK. The structure of a chondroitin sulfate-binding domain important in placental malaria. J Biol Chem. 2008;283:21842-6.
25. Singh K, Gittis AG, Nguyen P, Gowda DC, Miller LH, Garboczi DN. Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A. Nat Struct Mol Biol. 2008;15:932-8.
26. Singh SK, Hora R, Belrhali H, Chitnis CE, Sharma A. Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. Nature. 2006;439:741-4.
27. Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L. Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. Cell. 2005;122:183-93.
28. Shen H, Schuster R, Stringer KF, Waltz SE, Lentsch AB. The Duffy antigen/receptor for chemokines (DARC) regulates prostate tumor growth. FASEB J. 2006;20:59-64.
29. Pruenster M, Rot A. Throwing light on DARC. Biochem Soc Trans. 2006;34:1005-8.
30. Rot A. Chemokine patterning by glycosaminoglycans and interceptors. Front Biosci. 2010;15:645-60.

31. Borroni EM, Bonecchi R, Buracchi C, Savino B, Mantovani A, Locati M. Chemokine decoy receptors: new players in reproductive immunology. Immunol Invest. 2008;37:483-97.
32. Neote K, DiGregorio D, Mak JY, Horuk R, Schall TJ. Molecular cloning, functional expression, and signaling characteristics of a C-C chemokine receptor. Cell. 1993;72:415-25.
33. Hadley TJ, Lu ZH, Wasniowska K, Martin AW, Peiper SC, Hesselgesser J, et al. Postcapillary venule endothelial cells in kidney express a multispecific chemokine receptor that is structurally and functionally identical to the erythroid isoform, which is the Duffy blood group antigen. J Clin Invest. 1994;94:985-91.
34. Neote K, Mak JY, Kolakowski LF, Jr., Schall TJ. Functional and biochemical analysis of the cloned Duffy antigen: identity with the red blood cell chemokine receptor. Blood. 1994;84:44-52.
35. Reutershan J, Harry B, Chang D, Bagby GJ, Ley K. DARC on RBC limits lung injury by balancing compartmental distribution of CXC chemokines. Eur J Immunol. 2009;39:1597-607.
36. Pruenster M, Mudde L, Bombosi P, Dimitrova S, Zsak M, Middleton J, et al. The Duffy antigen receptor for chemokines transports chemokines and supports their promigratory activity. Nat Immunol. 2009;10:101-8.
37. Smith E, McGettrick HM, Stone MA, Shaw JS, Middleton J, Nash GB, et al. Duffy antigen receptor for chemokines and CXCL5 are essential for the recruitment of neutrophils in a multicellular model of rheumatoid arthritis synovium. Arthritis Rheum. 2008;58:1968-73.
38. Zarbock A, Schmolke M, Bockhorn SG, Scharte M, Buschmann K, Ley K, et al. The Duffy antigen receptor for chemokines in acute renal failure: A facilitator of renal chemokine presentation. Crit Care Med. 2007;35:2156-63.
39. Chakera A, Seeber RM, John AE, Eidne KA, Greaves DR. The duffy antigen/receptor for chemokines exists in an oligomeric form in living cells and functionally antagonizes CCR5 signaling through hetero-oligomerization. Mol Pharmacol. 2008;73:1362-70.
40. Lachgar A, Jaureguiberry G, Le Buenac H, Bizzini B, Zagury JF, Rappaport J, et al. Binding of HIV-1 to RBCs involves the Duffy antigen receptors for chemokines (DARC). Biomed Pharmacother. 1998;52:436-9.
41. He W, Neil S, Kulkarni H, Wright E, Agan BK, Marconi VC, et al. Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. Cell Host Microbe. 2008;4:52-62.
42. Beck Z, Brown BK, Wieczorek L, Peachman KK, Matyas GR, Polonis VR, et al. Human erythrocytes selectively bind and enrich infectious HIV-1 virions. PLoS One. 2009;4:e8297.
43. Kulkarni H, Marconi VC, He W, Landrum ML, Okulicz JF, Delmar J, et al. The Duffy-null state is associated with a survival advantage in leukopenic HIV-infected persons of African ancestry. Blood. 2009;114:2783-92.
44. Winkler CA, An P, Johnson R, Nelson GW, Kirk G. Expression of Duffy antigen receptor for chemokines (DARC) has no effect on HIV-1 acquisition or progression to AIDS in African Americans. Cell Host Microbe. 2009;5:411-3; author reply 8-9.
45. Julg B, Reddy S, van der Stok M, Kulkarni S, Qi Y, Bass S, et al. Lack of Duffy antigen receptor for chemokines: no influence on HIV disease

progression in an African treatment-naive population. Cell Host Microbe. 2009;5:413-5; author reply 8-9.
46. Walley NM, Julg B, Dickson SP, Fellay J, Ge D, Walker BD, et al. The Duffy antigen receptor for chemokines null promoter variant does not influence HIV-1 acquisition or disease progression. Cell Host Microbe. 2009;5:408-10; author reply 18-9.
47. Horton LW, Yu Y, Zaja-Milatovic S, Strieter RM, Richmond A. Opposing roles of murine duffy antigen receptor for chemokine and murine CXC chemokine receptor-2 receptors in murine melanoma tumor growth. Cancer Res. 2007;67:9791-9.
48. Zerfaoui M, Naura AS, Errami Y, Hans CP, Rezk BM, Park J, et al. Effects of PARP-1 deficiency on airway inflammatory cell recruitment in response to LPS or TNF: differential effects on CXCR2 ligands and Duffy Antigen Receptor for Chemokines. J Leukoc Biol. 2009;86:1385-92.
49. Lentsch AB. The Duffy antigen/receptor for chemokines (DARC) and prostate cancer. A role as clear as black and white? FASEB J. 2002;16:1093-5.
50. Wang J, Ou ZL, Hou YF, Luo JM, Shen ZZ, Ding J, et al. Enhanced expression of Duffy antigen receptor for chemokines by breast cancer cells attenuates growth and metastasis potential. Oncogene. 2006;25:7201-11.
51. Le Van Kim C, Tournamille C, Kroviarski Y, Cartron JP, Colin Y. The 1.35-kb and 7.5-kb Duffy mRNA isoforms are differently regulated in various regions of brain, differ by the length of their 5' untranslated sequence, but encode the same polypeptide. Blood. 1997;90:2851-3.
52. Tournamille C. [Molecular basis and structure-activity relationships of the Duffy blood group antigens: chemokine and Plasmodium vivax receptors]. Transfus Clin Biol. 2000;7:497-509.
53. Tournamille C, Blancher A, Le Van Kim C, Gane P, Apoil PA, Nakamoto W, et al. Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines. Immunogenetics. 2004;55:682-94.
54. Tournamille C, Le Van Kim C, Gane P, Blanchard D, Proudfoot AE, Cartron JP, et al. Close association of the first and fourth extracellular domains of the Duffy antigen/receptor for chemokines by a disulfide bond is required for ligand binding. J Biol Chem. 1997;272:16274-80.
55. Tournamille C, Le Van Kim C, Gane P, Le Pennec PY, Roubinet F, Babinet J, et al. Arg89Cys substitution results in very low membrane expression of the Duffy antigen/receptor for chemokines in Fy(x) individuals. Blood. 1998;92:2147-56.
56. Czerwinski M, Kern J, Grodecka M, Paprocka M, Krop-Watorek A, Wasniowska K. Mutational analysis of the N-glycosylation sites of Duffy antigen/receptor for chemokines. Biochem Biophys Res Commun. 2007;356:816-21.
57. Wasniowska K, Blanchard D, Janvier D, Wang ZX, Peiper SC, Hadley TJ, et al. Identification of the Fy6 epitope recognized by two monoclonal antibodies in the N-terminal extracellular portion of the Duffy antigen receptor for chemokines. Mol Immunol. 1996;33:917-23.
58. Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum. Science. 1994;264:1941-4.

59. Tournamille C, Filipe A, Badaut C, Riottot MM, Longacre S, Cartron JP, et al. Fine mapping of the Duffy antigen binding site for the Plasmodium vivax Duffy-binding protein. Mol Biochem Parasitol. 2005;144:100-3.
60. de Brevern AG. New opportunities to fight against infectious diseases and to identify pertinent drug targets with novel methodologies. Infect Disord Drug Targets. 2009;9:246-7.
61. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28:235-42.
62. Tusnady GE, Dosztanyi Z, Simon I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res. 2005;33:D275-8.
63. von Heijne G. Membrane-protein topology. Nat Rev Mol Cell Biol. 2006;7:909-18.
64. McLuskey K, Roszak AW, Zhu Y, Isaacs NW. Crystal structures of all-alpha type membrane proteins. Eur Biophys J. 2010;39:723-55.
65. Neumann S, Fuchs A, Mulkidjanian A, Frishman D. Current status of membrane protein structure classification. Proteins. 2010;78:1760-73.
66. White SH. The progress of membrane protein structure determination. Protein Sci. 2004;13:1948-9.
67. White SH. Biophysical dissection of membrane proteins. Nature. 2009;459:344-6.
68. Newstead S, Ferrandon S, Iwata S. Rationalizing alpha-helical membrane protein crystallization. Protein Sci. 2008;17:466-72.
69. Harrington SE, Ben-Tal N. Structural determinants of transmembrane helical proteins. Structure. 2009;17:1092-103.
70. de Brevern AG. 3D-structural models of transmembrane proteins. In: J.J L, editor. Membrane protein structure determination: from structure to function in Methods in Molecular Biology collection NJ, USA: Humana Press Inc.; 2010. p. in press.
71. Fernandez-Fuentes N, Madrid-Aliste CJ, Rai BK, Fajardo JE, Fiser A. M4T: a comparative protein structure modeling server. Nucleic Acids Res. 2007;35:W363-8.
72. Kosinski J, Gajda MJ, Cymerman IA, Kurowski MA, Pawlowski M, Boniecki M, et al. FRankenstein becomes a cyborg: the automatic recombination and realignment of fold recognition models in CASP6. Proteins. 2005;61 Suppl 7:106-13.
73. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. Crystal structure of rhodopsin: A G protein-coupled receptor. Science. 2000;289:739-45.
74. Wasniowska K, Petit-LeRoux Y, Tournamille C, Le van Kim C, Cartron JP, Colin Y, et al. Structural characterization of the epitope recognized by the new anti-Fy6 monoclonal antibody NaM 185-2C3. Transfus Med. 2002;12:205-11.
75. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng. 1997;10:673-6.
76. von Heijne G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol. 1992;225:487-94.
77. Claros MG, von Heijne G. TopPred II: an improved software for membrane protein structure predictions. Comput Appl Biosci. 1994;10:685-6.

78. Tusnady GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol. 1998;283:489-506.
79. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001;17:849-50.
80. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567-80.
81. Rost B, Casadio R, Fariselli P, Sander C. Transmembrane helices predicted at 95% accuracy. Protein Sci. 1995;4:521-33.
82. Rost B, Fariselli P, Casadio R. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci. 1996;5:1704-18.
83. Hofmann K, Stoffel W. TMbase - A database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler. 1993;374:166-9.
84. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics. 1998;14:378-9.
85. Mitaku S, Hirokawa T. Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. Protein Eng. 1999;12:953-7.
86. Mitaku S, Hirokawa T, Tsuji T. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. Bioinformatics. 2002;18:608-16.
87. Juretic D, Lee B, Trinajstic N, Williams RW. Conformational preference functions for predicting helices in membrane proteins. Biopolymers. 1993;33:255-73.
88. Juretic D, Zoranic L, Zucic D. Basic charge clusters and predictions of membrane protein topology. J Chem Inf Comput Sci. 2002;42:620-32.
89. Pasquier C, Hamodrakas SJ. An hierarchical artificial neural network system for the classification of transmembrane proteins. Protein Eng. 1999;12:631-4.
90. Pasquier C, Promponas VJ, Palaios GA, Hamodrakas JS, Hamodrakas SJ. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. Protein Eng. 1999;12:381-5.
91. Persson B, Argos P. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J Mol Biol. 1994;237:182-92.
92. Persson B, Argos P. Topology prediction of membrane proteins. Protein Sci. 1996;5:363-71.
93. Kihara D, Shimizu T, Kanehisa M. Prediction of membrane proteins based on classification of transmembrane segments. Protein Eng. 1998;11:961-70.
94. Deber CM, Wang C, Liu LP, Prior AS, Agrawal S, Muskat BL, et al. TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. Protein Sci. 2001;10:212-9.
95. Zhou H, Zhou Y. Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model-based method. Protein Sci. 2003;12:1547-55.
96. Jones DT, Taylor WR, Thornton JM. A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry. 1994;33:3038-49.

97. Jones DT. Do transmembrane protein superfolds exist? FEBS Lett. 1998;423:281-5.
98. Viklund H, Elofsson A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. Protein Sci. 2004;13:1908-17.
99. Shen H, Chou JJ. MemBrain: improving the accuracy of predicting transmembrane helices. PLoS ONE. 2008;3:e2399.
100. Cao B, Porollo A, Adamczak R, Jarrell M, Meller J. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. Bioinformatics. 2006;22:303-9.
101. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292:195-202.
102. Pylouster J, Bornot A, Etchebest C, de Brevern AG. Influence of assignment on the prediction of transmembrane helices in protein structures. Amino Acids. 2010.
103. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577-637.
104. de Brevern AG, Etchebest C, Hazout S. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins. 2000;41:271-87.
105. Etchebest C, Benros C, Hazout S, de Brevern AG. A structural alphabet for local protein structures: improved prediction methods. Proteins. 2005;59:810-27.
106. Sali A, Blundell TL. Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. J Mol Biol. 1990;212:403-28.
107. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234:779-815.
108. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct. 2000;29:291-325.
109. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci. 2003;12:2001-14.
110. Kutzner C, van der Spoel D, Fechner M, Lindahl E, Schmitt UW, de Groot BL, et al. Speeding up parallel GROMACS on high-latency networks. J Comput Chem. 2007;28:2075-84.
111. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26:1701-18.
112. Hubbard SJ, Thornton JM. 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London.; 1993.
113. DeLano WLT. The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA http://wwwpymolorg. 2002.
114. Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res. 2007;35:3375-82.
115. Wu S, Zhang Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins. 2008;72:547-56.

116. Gao X, Bu D, Xu J, Li M. Improving consensus contact prediction via server correlation reduction. BMC Struct Biol. 2009;9:28.
117. Guo JT, Ellrott K, Chung WJ, Xu D, Passovets S, Xu Y. PROSPECT-PSPP: an automatic computational pipeline for protein structure prediction. Nucleic Acids Res. 2004;32:W522-5.
118. Zhou H, Skolnick J. Protein structure prediction by pro-Sp3-TASSER. Biophys J. 2009;96:2119-27.
119. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005;21:951-60.
120. Zhou H, Zhou Y. SPARKS 2 and SP3 servers in CASP6. Proteins. 2005;61 Suppl 7:152-6.
121. Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol. 2001;310:243-57.
122. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009;4:363-71.
123. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, et al. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science. 2008;322:1211-7.
124. Hoffman GR, Rahl PB, Collins RN, Cerione RA. Conserved structural motifs in intracellular trafficking pathways: structure of the gammaCOP appendage domain. Mol Cell. 2003;12:615-25.
125. Dias FM, Vincent F, Pell G, Prates JA, Centeno MS, Tailford LE, et al. Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of Cellvibrio mixtus mannosidase 5A. J Biol Chem. 2004;279:25517-26.
126. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. J Mol Biol. 2004;342:571-83.
127. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A messagepassing parallel molecular dynamics implementation. Comp Phys Comm. 1995;91:43-56.
128. de Brevern AG. New assessment of a structural alphabet. In Silico Biol. 2005;5:283-9.
129. de Brevern AG, Etchebest C, Benros C, Hazout S. "Pinning strategy": a novel approach for predicting the backbone structure in terms of protein blocks from sequence. J Biosci. 2007;32:51-70.
130. de Brevern AG, Benros C, Gautier R, Valadie H, Hazout S, Etchebest C. Local backbone structure prediction of proteins. In Silico Biol. 2004;4:381-6.
131. Hollup SM, Salensminde G, Reuter N. WEBnm@: a web application for normal mode analyses of proteins. BMC Bioinformatics. 2005;6:52.
132. Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res. 2006;34:W52-6.
133. Suhre K, Sanejouand YH. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. 2004;32:W610-4.
134. Suhre K, Sanejouand YH. On the potential of normal-mode analysis for solving difficult molecular-replacement problems. Acta Crystallogr D Biol Crystallogr. 2004;60:796-9.

135. Rocchia W, Alexov B, Honig B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. J Phys Chem B. 2001;105 6507-14.
136. Crump MP, Gong JH, Loetscher P, Rajarathnam K, Amara A, Arenzana-Seisdedos F, et al. Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. EMBO J. 1997;16:6996-7007.
137. Baysal C, Atilgan AR. Elucidating the structural mechanisms for biological activity of the chemokine family. Proteins. 2001;43:150-60.
138. Fernandez-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins. 2003;52:113-7.
139. ICM. http://www.molsoft.com/.
140. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res. 2004;32:W96-9.
141. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics. 2004;20:45-50.
142. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa EB, et al. Naturally occurring antibodies devoid of light chains. Nature. 1993;363:446-8.
143. Muyldermans S, Baral TN, Retamozzo VC, De Baetselier P, De Genst E, Kinne J, et al. Camelid immunoglobulins and nanobody technology. Vet Immunol Immunopathol. 2009;128:178-83.
144. Baral TN, Magez S, Stijlemans B, Conrath K, Vanhollebeke B, Pays E, et al. Experimental therapy of African trypanosomiasis with a nanobody-conjugated human trypanolytic factor. Nat Med. 2006;12:580-4.
145. Abderrazek RB, Hmila I, Vincke C, Benlasfar Z, Pellis M, Dabbek H, et al. Identification of potent nanobodies to neutralize the most poisonous polypeptide from scorpion venom. Biochem J. 2009;424:263-72.
146. Dumoulin M, Last AM, Desmyter A, Decanniere K, Canet D, Larsson G, et al. A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. Nature. 2003;424:783-8.
147. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-402.
148. De Genst E, Handelberg F, Van Meirhaeghe A, Vynck S, Loris R, Wyns L, et al. Chemical basis for the affinity maturation of a camel single domain antibody. J Biol Chem. 2004;279:53593-601.
149. Wootton JC, Federhen S. Analysis of compositionally biased regions in sequence databases. Methods Enzymol. 1996;266:554-71.
150. Decanniere K, Transue TR, Desmyter A, Maes D, Muyldermans S, Wyns L. Degenerate interfaces in antigen-antibody complexes. J Mol Biol. 2001;313:473-8.
151. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-80.
152. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst. 1993;26:283-91.