To be published in Infectious Diseases – Drug Targets (2009)

in silico studies on DARC

Alexandre G. de Brevern*, Ludovic Autin[#], Yves Colin, Olivier Bertrand & Catherine Etchebest

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

Short title: in silico DARC

* Corresponding author:
mailing address: Dr. de Brevern A.G., INSERM UMR-S 665, Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), Institut National de la Transfusion Sanguine (INTS), 6, rue Alexandre Cabanel, 75739 Paris 15, France
E-mail: alexandre.debrevern@univ-paris-diderot.fr
Tel: (33) 1 44 49 30 38
Fax: (33) 1 43 06 50 19

[#] present adress : Molecular Graphics Laboratory, Department of Molecular Biology, MB-5, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037-1000, USA.

Abstract

The Duffy Antigen/Receptor for Chemokine (DARC) is a seven segment transmembrane protein. It was firstly discovered as a blood group antigen and was the first specific gene locus assigned to a specific autosome in man. It became more famous as an erythrocyte receptor for malaria parasites (*Plasmodium vivax* and *Plasmodium knowlesi*), and finally for chemokines. DARC is an unorthodox chemokine receptor as (i) it binds chemokines of both CC and CXC classes and (ii) it lacks the Asp-Arg-Tyr consensus motif in its second cytoplasmic loop hence cannot couple to G proteins and activate their signaling pathways. DARC had also been associated to cancer progression, numerous inflammatory diseases, and possibly to AIDS.

In this review, we will summarize important biological data on DARC. Then we shall focus on recent development of the elaboration and analyzes of structural models of DARC. We underline the difficulty to propose pertinent structural models of transmembrane protein using comparative modeling process, and other dedicated approaches as the Protein Blocks. The chosen structural models encompass most of the biochemical data known to date. Finally, we present recent development of protein – protein docking between DARC structural models and CXCL-8 structures. We propose a hierarchal search based on separated rigid and flexible docking.

Key-words: Duffy Antigen / Receptor for Chemokine; CXCL-8, Duffy Binding Protein, chemokines, malaria, *Plasmodium vivax*.

Biology and Pathology of the DARC

The Duffy antigen receptor for chemokines (DARC) is a protein which has a long history. It was discovered at the beginning of the 50's as a blood group antigen, after the identification of antibodies against the protein in the serum of a polytransfused haemophilic named Duffy. DARC carries the antithetic antigens Fy^a and Fy^b in humans [1, 2]. DARC, as a blood group, was extensively studied and its mode of inheritance defined [3-12]. Interestingly, it was the first specific gene locus assigned to a specific autosome in man [13].

A milestone in DARC studies was the discovery that it is an erythrocyte receptor for malaria parasites. Indeed, *in vitro* and *in vivo* studies performed on American volunteering detainees have shown that erythrocytes of $Fy(a^-b^-)$ individuals that do not express Duffy cannot be invaded by *Plasmodium knowlesi* [14] and *Plasmodium vivax* [15]. The Duffy negative phenotype is due to homozygosity for a promoter polymorphism (-46C) that disrupts the binding site for a transcription factor (GATA-1) required for the DARC chemokine receptor to be expressed on the cell surface of erythrocytes [16]. This mutation is ubiquitous in West Africa proving that it was due to a strong selection pressure. An important point to be stressed is that this mutation abolishes expression of DARC in the erythrocytic lineage but does not hinder expression in other cell types in which another promoter is operative.

It must be noticed that *Plasmodium vivax* malaria [17] is the most geographically widespread and the second prevalent cause of malaria. Among 2.6 billion people at risk of malaria infection, annual estimates of *P. vivax* cases range from 130 to 435 million [18]. 90% of these infections occur outside Africa [19]. *P. vivax* transmission agents to humans are *Anopheles* of various species [20]. Chloroquine was one of the first and still the most commonly used drug for treatment for *P. vivax* malaria. However, chloroquine resistance *P. vivax* becomes a threatening problem [21-29].

Searching for a vaccine eliciting antibodies able to interfere with interaction of DARC

with its malarial partner the Duffy Binding Protein (or DBP) is an active research field [30-33]**.** However some recent data were issued in the literature suggesting that *P. vivax* might infest Duffy negative people in East Africa and Brazil [34-37]

DARC, while it was well known as a receptor for *P. vivax* was more recently identified as a chemokine receptor because Duffy-positive but not Duffy-null erythrocytes can bind CXCL8 (IL-8) [38, 39] and is involved in physiological chemiotactism [40]. Contrary to other chemokine receptors, it binds members of both the CC and CXC chemokine families. It binds angiogenic ELR^+ (*i.e.* with the Glutamine–Leucine–Arginine motif at the N-terminus) CXC chemokines, as well as some CC chemokines [41-43], but it does not bind ELR^- CXC chemokines and the C chemokine [44]. DARC ligands with high affinity include CXCL1 (GROα), CXCL2 (GROβ), CXCL3 (GROγ), CXCL4 (PF4), CXCL5 (ENA-78), CXCL7 (NAP-2), CXCL8 (IL-8), CCL2 (MCP-1, MCAF, JE), CCL5 (RANTES), CCL7 (MCP-3), CCL11 (Eotaxin), CCL13 (MCP-4), CCL14 (HCC-1) and CCL17 (TARC) [45].

The DARC receptor plays a complex role in regulating chemokine levels, at least in part due to its ability to bind chemokines and adsorb them onto the red cell surface, thereby acting as a chemokine reservoir; it is so considered as typical decoy receptor or a sink [38]. Chemokines binding to DARC can largely be categorized as inflammatory. Chemokine binding to DARC does not lead to intracellular signaling. Because DARC lacks the Asp-Arg-Tyr consensus motif in its second cytoplasmic loop, it cannot couple to G proteins and activate the subsequent signaling pathways [46].

DARC was associated to numerous physiological and pathological responses [47] as the other atypical chemokine receptor D6 [48]. DARC is thought to play a part in homeostasis either through clearance of inflammatory chemokines from the circulation or maintenance of plasma levels by subsequent release from the red cell surface [49]. But the mechanisms by which DARC is implicated in inflammation are far to be completely understood. It has been

reported that DARC might heterodimerize with CXCR-5 and so interfere with function of this important effector [50].

DARC expressed in endothelial cells of post capillary venules might assist leukocyte migration to sites of inflammation [46], Noticeably it has been suggested that DARC might ensure transcytosis of chemokines from the intravascular to the extravascular spaces hence favor migration of leucocytes involved in inflammation [51]. This mechanism clearly is operative even in individuals lacking Duffy antigen expression on red cells. Interactions of chemokines with DARC support their activity on apposing leukocytes *in vitro* and *in vivo* [49]. The up-regulation of DARC expression by endothelial cells has been demonstrated in many human inflammatory and infectious diseases [52-59]. For instance, DARC is implicated in numerous pathologies implicating "classical" chemokine receptors like in giant cell arteritis [52]. Besides its role in inflammation, definitely DARC seems to play a role in cancer. Epidemiological data suggest that lack of DARC on erythrocytes favors early onset and aggressiveness of prostate cancer observed in African Americans [60]. These data are backed up by experimental models using genetically modified mice [61]. On the other hand expression of DARC on malignant lung or breast cancer cells might be associated with less growth and metastatic potential this being possibly related to sequestration of angiogenic chemokines and inhibition of tumor vascularisation [62, 63]. Susceptibility to asthma and atopy among certain populations of African descent was correlated also to absence of DARC expression on red cells [64].

A recent study suggests that DARC influences HIV/AIDS susceptibility by mediating trans-infection of HIV-1 and by affecting both chemokine-HIV interactions and chemokine-driven inflammation [65]. The almost universal presence of the HIV-susceptible DARC genotype in Africans might be an important contribution to understanding the massive extent of the HIV-1 epidemic in Africa [66, 67]. Worth of notice This results was considered as one

of the eight most important discovery of the year in human and medical genetics by the Annals of New York Academy of Sciences [68].

Expression in cell lines of mutant DARC molecules and studies of their affinity for CXCL8 and various antibodies [69-79] helped to define the coarse topology of DARC [80], *i.e.*, a 7 transmembrane protein with four extracellular loops (named Extra Cellular Domains or *ECDs*) and four intracellular loops (named Intra Cellular Domains or *ICDs*). These experimental data coupled with sophisticated molecular modeling had allowed proposing pertinent structural models [81] able to explain the docking of DARC with CXCL-8 [82] and in the future hopefully with the malaria protein DBP.

Principle of the *in silico* approach

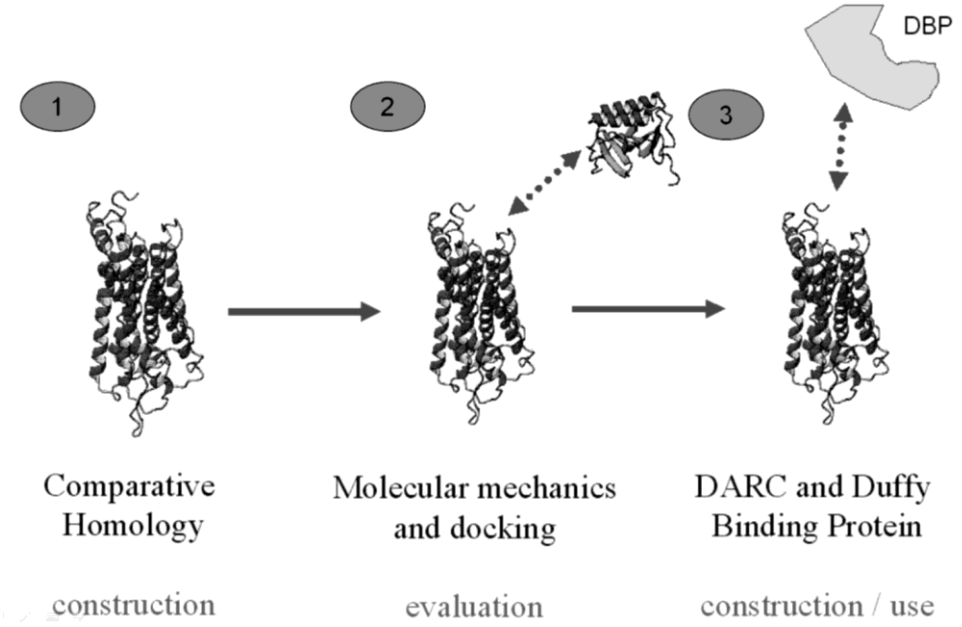

Figure 1. The three steps. (1) by comparative modeling, structural models are building; (2) docking approach is performed with the natural ligand structure, *i.e.*, CXC-L8 and (3) with the Duffy Binding Protein (DBP).

Transmembrane protein represents about ~25% of proteins coded by genomes. They are composed of two major classes: all-α, *e.g.* rhodopsin and all-β, *e.g.*, Outer Membrane Proteins. They are the support of essential biological functions acting as receptors, transporters or channels. Moreover, 2/3 of the marketed drugs targets a transmembrane

protein and 50% specifically a GPCR [83], obtaining atomic structures becomes a major axis of research. Yet, these proteins are embedded in a lipid membrane that constitutes a very specific environment. Because the proteins are strongly destabilized when extracted from their natural medium, 3D transmembrane structures are difficult to obtain experimentally. On that account, the total number of transmembrane proteins in the Protein DataBank [84] is limited, representing ~1% of all the available structures [85]. Thus alternative approaches are required to obtain structural information. Consequently methods aiming at constructing 3D structural models are become an important research fields for understanding biological mechanisms and interactions [86].

Figure 1 shows the principle of the *in silico* approach adopted to build structural DARC models and to dock its ligands. It can be roughly divided into three steps. Firstly, pertinent structural models were built using comparative modeling. Then, after refinement of different models, docking of the DARC natural ligand, *i.e*. CXC-L8, was performed by sophisticated docking techniques using the most relevant structural models. Finally, similar docking approaches can be applied to the malarial protein DBP.

The structural models

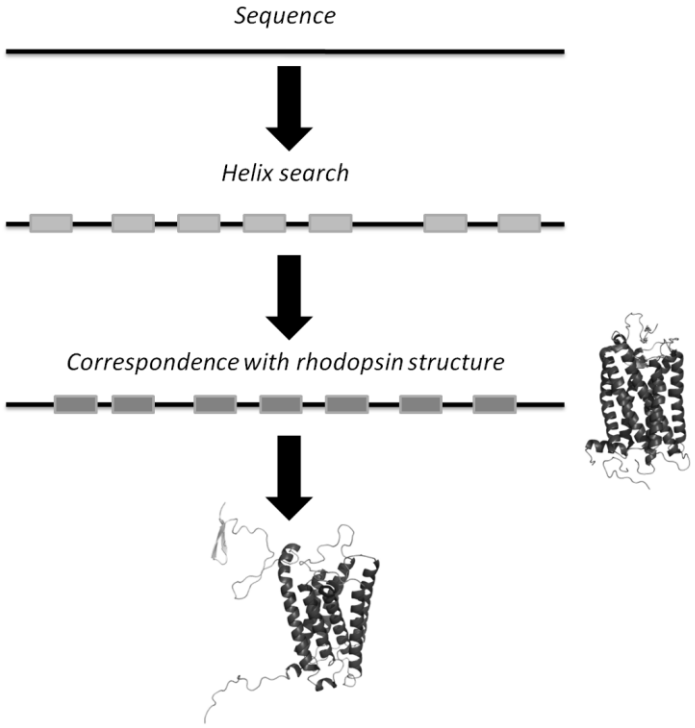

Figure 2. Principles of homology modeling for a transmembrane protein.

Building a structural model of a transmembrane protein related to GPCR can be decomposed in simple tasks (cf. Figure 2 and [87]). Once the helical transmembrane regions are located with dedicated prediction software, these predicted helices are aligned with the corresponding TM helices of available GPCR structures, *e.g.* rhodopsin (PDB code 1F88 [88]). However few transmembrane proteins can easily fit into this idealistic pipeline. For two main reasons: (i) the sequence similarity with rhodopsin (as well with beta2-adrenergic receptor as A2A adenosine receptor) is often low and (ii) prediction of transmembrane segments requires special attention and caution.

From experimental evidence, DARC protein can be divided into three main domains: (i) one central domain encompassing the transmembrane helices and connection loops, (ii) the long N terminus region, namely *ECD1* and (iii) the short C terminus region, namely *ICD4*.

For none of these domains, neither PSI-BLAST [89] nor FASTA [90], the most known sequence similarity search tools, retrieve protein structures related to DARC in the Protein

Data Bank Moreover DARC family has only at most 25% of identity with CXCR-2 family. Phylogenetic analyses show that DARC family is highly conserved but very different from other chemokine receptor sequences (see Figure 3, [71, 91-93]).

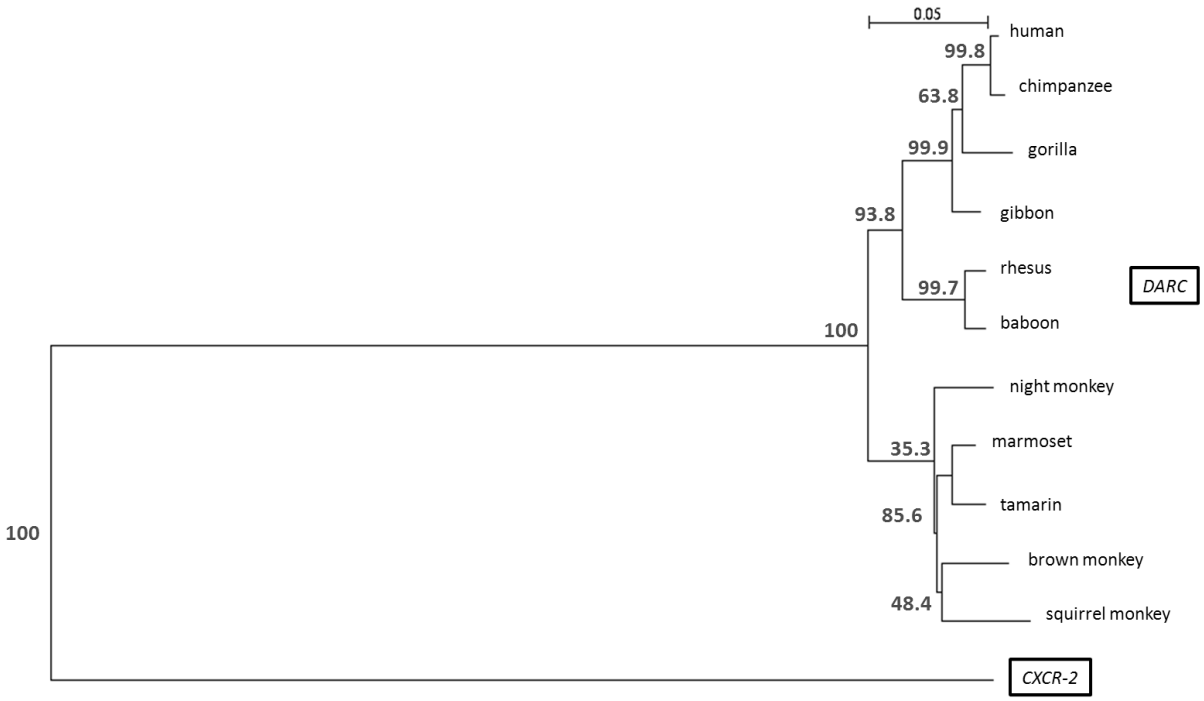

Figure 3. Phylogenetic analysis of DARC. Different sequences of DARC were aligned using CLUSTALW [91].The dendogram is plotted with Njplot software [92]. Bootstrap values computed by CLUSTALW are shown at the different nodes. Sequences were extracted from UniProt/UniParc ID [94]: the major 336 aa [93] human (DUFF_HUMAN, Q16570 [95]) was used, gorilla (*Gorilla gorilla*), AF311914; marmoset (*Callithrix jacchus*), AF311915; tamarin (*Saguinus oedipus*), AF311916; night monkey (*Aotus trivirgatus*), AF311917; squirrel monkey (*Samiri boliviensis*), AF311918; brown capuchin (*Cebus apella*), AF311919; chimpanzee (*Pan troglodytes*), AF311920; rhesus monkey (*Macaca mulatta*), AF311921; baboon (*Papio papio*), AF303532; gibbon (*Hylobates lar*) AF303533 (all the last sequences are from [71]).

Surprisingly; when DARC fragments are considered instead of the complete protein, only short similar fragments mainly related to globular β-sheets are found in the PDB. These results are in conflict with most of the data available indicating that the central domain of DARC receptor is mainly composed of α-helices. Actually, alignment of DARC and rhodopsin is exactly at the level of a random alignment (12%).

The prediction of transmembrane segments is also quite a complex problem. Numerous methods are at present available. They differ on the learning data used and on the kind of

learning and prediction algorithms. For instance, HMMTOP [96, 97] and TMHMM [98] are based on hidden Markov models. Yet, they give different results, even if the expected prediction rate –in regards to the literature- is supposed to be quite impressive. We have tested a dozen of these methods: DAS [99], TopPred 2 [100, 101], HMMTOP [96, 97], TMHMM 1.0 and 2.0 [98], PHDtm [102, 103], TMpred [104], SOSUI [105-107], SPLIT [108, 109], Pred-TMR 1.0 and 2.0 [110, 111], TMAP [112, 113], TSEG [114], TM-FINDER [115], UMDHMM^{TMHP} [116], MEMSAT [117, 118], PRODIV-TMHMM [119] and MemBrain [120]. Figure 4 summarizes the results obtained with most of them for DARC sequence [121]. Only the first helix is predicted with a large consensus while the fourth helix has a huge range of definitions, diverging sometimes by 15 residues.

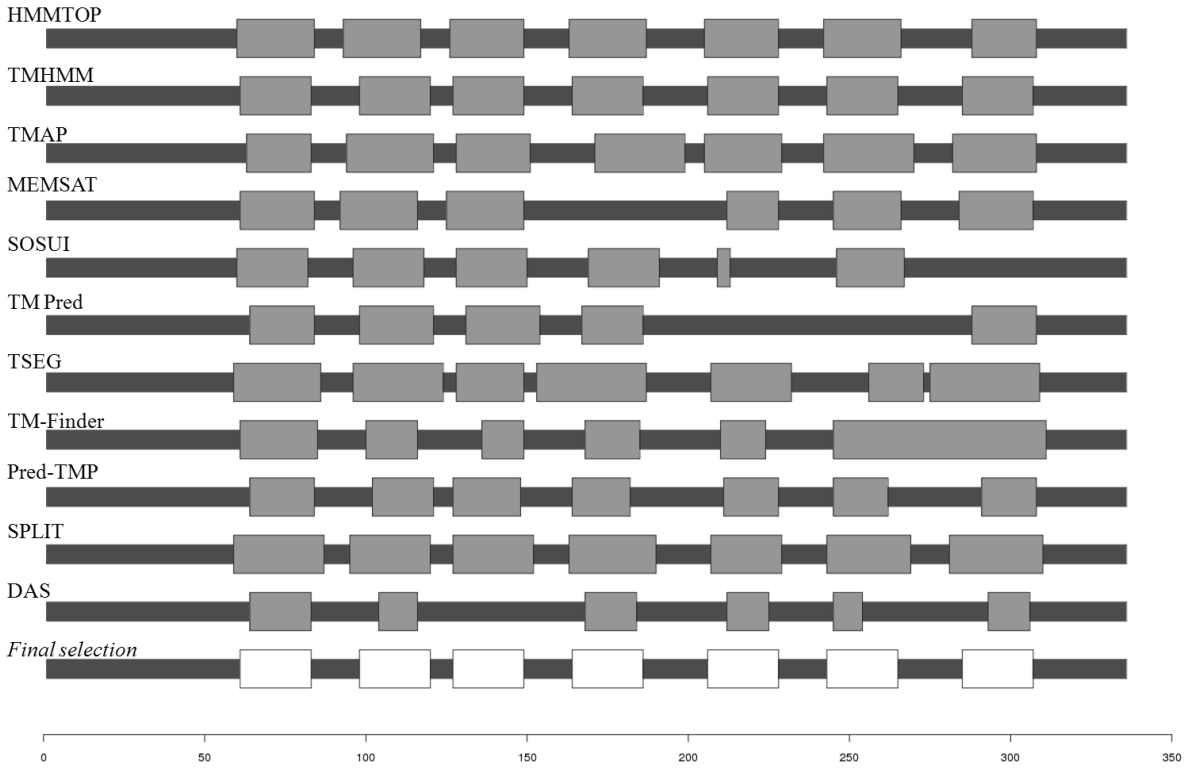

Figure 4. Prediction of transmembrane segments of DARC. Are shown the predictions done using HMMTOP [96, 97], TMHMM 2.0 [98], TMAP [113], MEMSAT [117], SOSUI [107], TMpred [104], TSEG [114], TM-FINDER [115], Pred-TMR 2.0 [111], SPLIT [109] and DAS [99]. The final selection corresponds to the positions finally obtained after multiple tests and evaluations. The plot was done using R software [121].

So, we designed a supervised approach to obtain pertinent structural models. This

approach benefited from a crucial element, *i.e.* the availability of experimental data that guided the whole procedure. These data consisted in the measure of the CXCL8 binding rate with 40 mutants. These measures underline the potential accessibility of some residues. Without these experimental data, assessment of the resulting structural models would be extremely difficult. Data were collected from the literature and our lab [71, 72, 122]. Figure 3 describes the methodology we have used to obtain pertinent structural models of DARC.

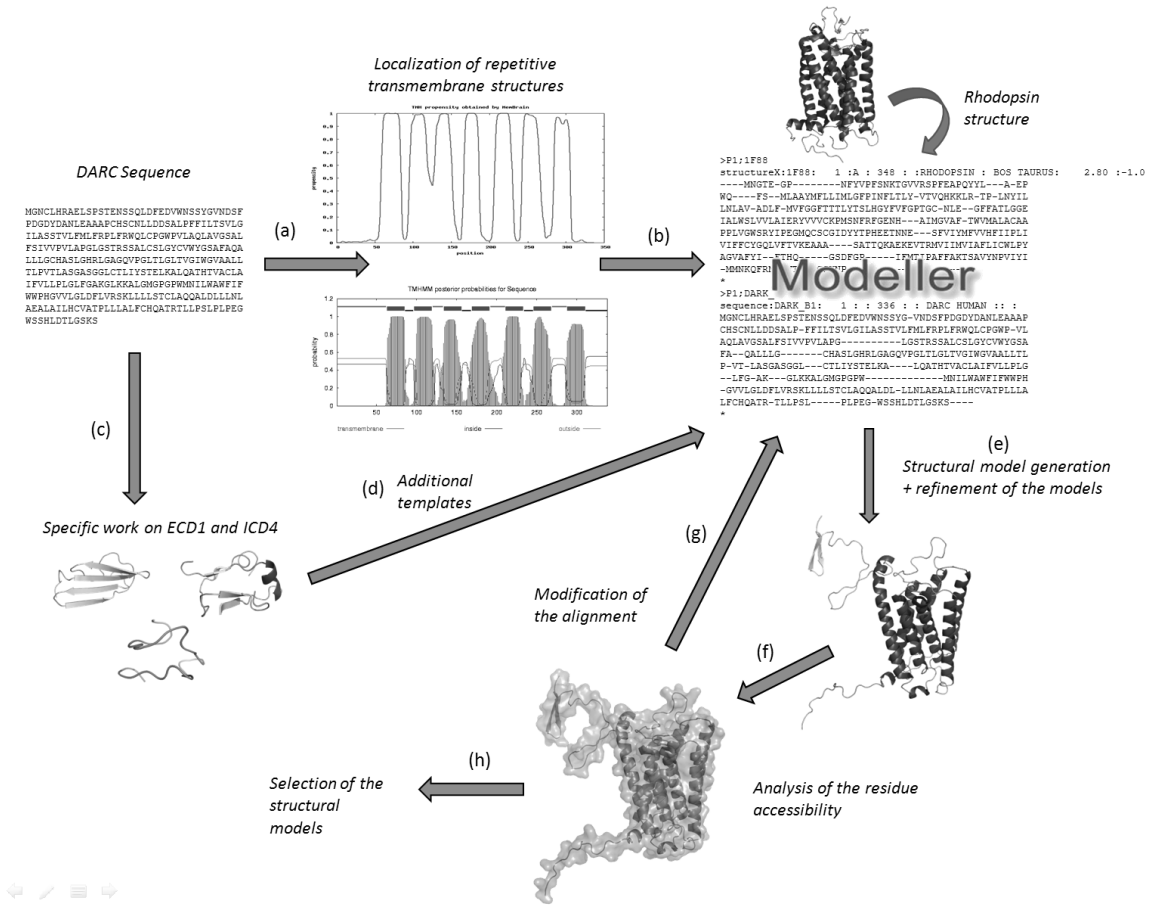

Figure 5. Building structural models of DARC. (a) Prediction of transmembrane helices. (b) Alignment of helical regions with corresponding regions of rhodopsin structure. (c) Potential structural templates for .*ECD1* and *ICD4* prediction of (d) Addition of these results to the complete alignment for comparative modeling. (e) Structural model generation and refinement of these models. (f) Accessibility computation of amino acids and known to be exposed. (g) In regards to the results, the alignment is modified. (h) At last, some models are selected.

A rough localization of the DARC transmembrane helical segments is done with different prediction algorithms presented in Figure 4 (cf. Figure 5a). The predicted α-helices of DARC are aligned to the assigned helices of the rhodopsin protein (PDB code: 1F88, chain

A [88]). Secondary structure assignment was done using DSSP software [123] (cf. Figure 5b). Similarly, predicted interconnecting loops of DARC were assigned systematically to their counterparts of rhodopsin. Potential structural homologues of *ECD1* and *ICD4* were searched with sophisticated approaches (Figure 5c) and the resulting propositions were added as additional templates (Figure 5d, see below).

The alignments with the sequences of the structural templates were done with Clustalw software [91] and manually modified. Hundred models were generated using Modeller software [124-126]. Each model was then refined, *e.g.*, using SCWRL [127] and GROMACS [128, 129] (cf. Figure 5e) and the residue accessibility was computed with Naccess software [130] (cf. Figure 5f). The alignment was modified in the light of the biological data known. Hence, different alternative locations of helical regions were tested and the corresponding alignments were modified accordingly (cf. Figure 5g). More than ten iterations of steps 5e to 5g were done and finally, models corresponding at best to the biological data were selected (cf. Figure 5h).

As notified in Figure 5c, we made a specific research for the longest extra-membranous domains of the protein, namely *ECD1* and *ICD4*. These two regions are exposed to solvent and consequently, they exhibit physico-chemical features similar to those of globular proteins. Secondary structure prediction methods, *e.g.* [131] predicted them as coil region with a medium confidence index. Turns prediction underlined many turns in both fragments [132]. HMMSTR, which is based on a more complex approach, predicted a short β-strand of 4-5 residues in *ECD1*, and a helical structure in *ICD4*, but the confidence index was poor [133].

Threading, *ab initio* and *de novo* methods were so required to find potential related protein fragments. The solutions were different but systematically, a high β-strand propensity was indicated in the first part of ECD1 and a helical structure at its end, close to the connection with the TM1 segment. In this research, we also used our own prediction approach

based on Protein Blocks.

Protein Blocks

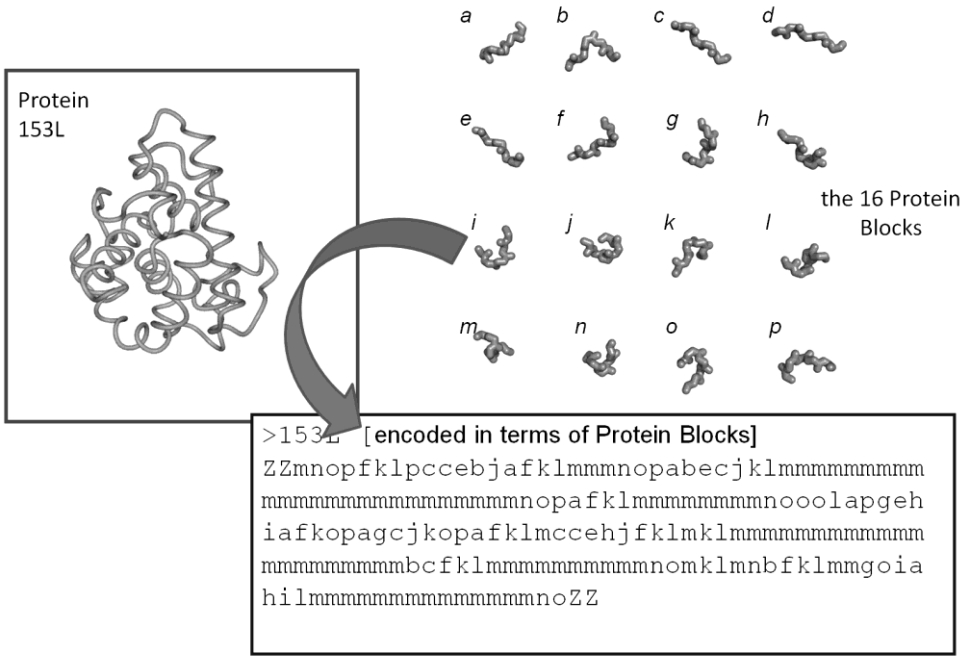

Figure 6. Encoding of the protein structures (3D) in terms of Protein Blocks.

The classical description of protein structures involves two regular states, the α-helices and the β-strands and one non-regular and variable state, the coil. Nonetheless, this simple definition of secondary structures masks numerous limitations. Indeed, three states may over-simplify the description of protein structure; 50% of all residues, *i.e.*, the coil, are not described even if it encompasses recurrent and similar local protein structures. Description of local protein structures have hence focused on the elaboration of complete sets of small prototypes or "structural alphabets" (SAs), that help to analyze local protein structures and to approximate every part of the protein backbone [134-137]. The principle of a structural alphabet firstly consists in designing a set of average recurrent local protein structures, able to approximate efficiently, every part of known structures (see Figure 6). Each residue being associated to one of these prototypes, then so, the 3D information of the protein structures can

be translated as a series of prototypes (letters) in 1D, as the amino acid sequence.

Our structural alphabet is composed of 16 structurally averaged protein fragments that are 5 residues in length, called Protein Blocks (PBs, [138]). They have been used both to describe the 3D protein backbones [139] and to perform a local structure prediction [138, 140-143]. Our works on PBs have proven their efficiency in the description and the prediction of long fragments [140, 144-147] and short loops [148], to define a reduced amino acid alphabet dedicated for mutation design [149], to analyze protein contacts [150] or in the building of a transmembrane protein [81]. We have also used protein blocks to compare / superimpose protein structures [151-153] (see Figure 1b).

Developments made in other laboratories, using PBs, have focused on the reconstruction of globular protein structures [154], the design of peptides [155], the definition of binding site signatures [156], novel prediction methodologies [157, 158] and fragment-based local statistical potentials [157]. The features of this alphabet have been compared by Karchin *et al*. [159] with those of 8 other structural alphabets showing that our PB alphabet is highly informative, with the best predictive ability of those tested. It is nowadays the most widely used SA in the world.

ECD1 prediction done with Protein Blocks method confirmed that the sequence is weakly informative. Nonetheless, it helped to distinguish 5 zones corresponding to (i) a first short region of extended structures, (ii) a second region related to extended structures following a short kink (iii) a third poorly informative region with a turn, confirmed by [132], (iv) a fourth fuzzy non-helical region (v) a fifth region predicted as mainly helical, with high confidence index, followed by a short loop.

Two selected models

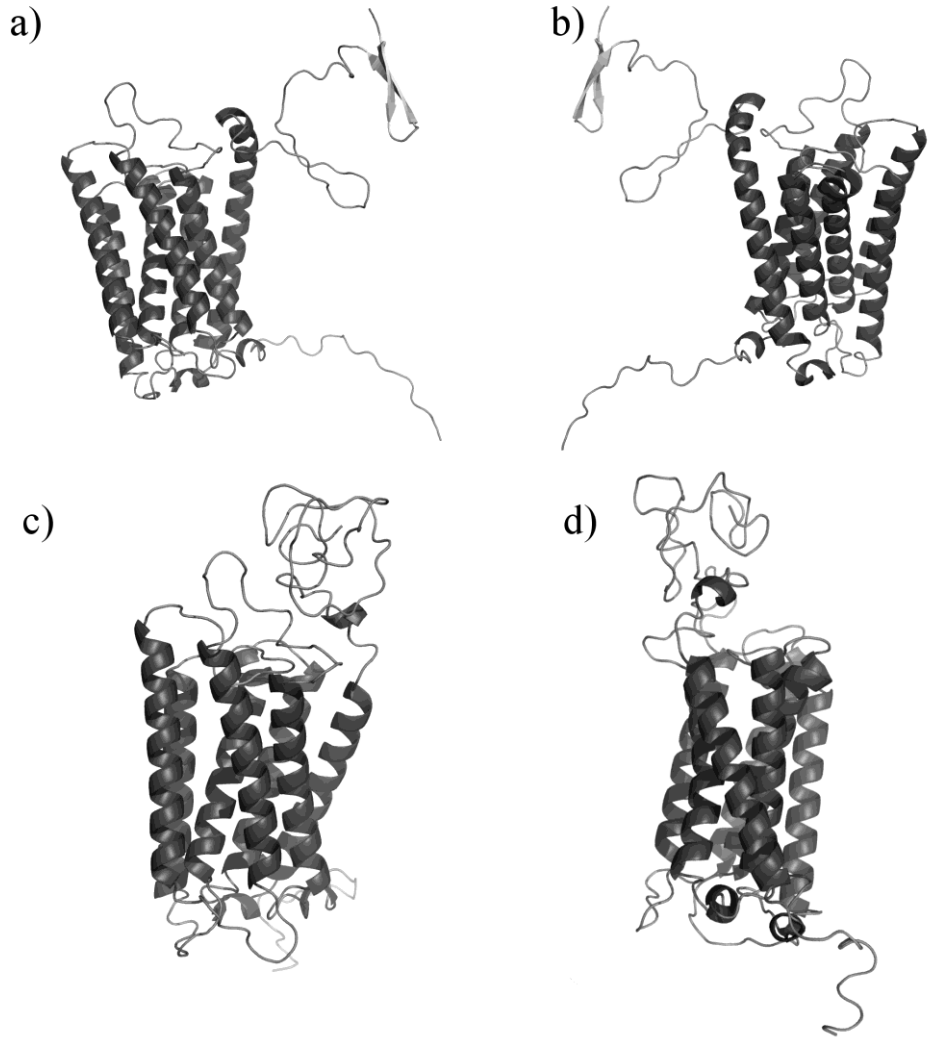

Figure 7. The two DARC selected models. Two views of (a-b) open form structural model and (c-d) closed structural model. Visualization done with PyMol software [160].

Finally, two models were selected (see Figure 7). Interestingly, the final positions of transmembrane helices gave a better sequence identity with rhodopsin structure than initially computed using classical alignment tools, *i.e.* 25% of sequence identity for the helices. The two models were quite different; in the first one, *ECD1* looked like a C letter and was far from the rest of the protein. This conformation was qualified as “open” (see Figures 7a and 7b). In contrast, in the second model, *ECD1* was more compact with a position closer to the rest of the protein. The conformation of the second model was denominated “closed” (see Figures 7c and 7d). Both models encompassed most of the 27 amino acids supposed to be

exposed. Exceptions were related to two amino acids found buried while not involved in disulfide bridges. The first one is a Phenylalanine (F65) located in the beginning of the first helix. The side chain was stacked with other aromatic amino acids. The second one is an Aspartate (D263) not close to the surface. These discrepancies may be related to the quality of the models but also to the crude interpretation we made of the experimental data.

To explore the flexibility of the *ECD* loops, we performed simulated annealing simulations [161]. Firstly, it explains well the importance of D263 which interacts directly with the end of the ECD3 . Secondly, when we analysed all the simulated structures with the use of Protein Blocks, we observed that some regions of *ECD1* tent to be more helical and other ones to be more extended. Interestingly, these regions corresponded to those predicted to be more extended or more helical with the prediction tools.

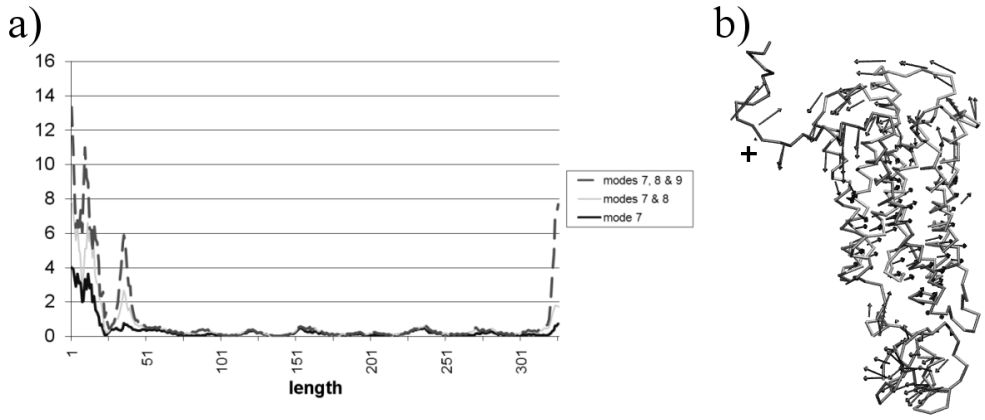

Figure 8. Normal Mode Analysis of DARC 'open' model. (a) (y-axis) Modes 7 to 9 for (x-axis) the whole sequence. (b) Visualization of the modes 7 to 9 on the 'open' model. The cross highlights the hinge region of *ECD1*. Visualization done with VMD software [162, 163].

Experimental data show that ECD1 is involved in the binding with ligands. The quite different position of ECD1 between the two models led us to perform a preliminary study aiming at analyzing domain motions. This study could lead to alternative positions of ECD1 domain. For that purpose, normal mode analysis is well adapted. We use the WEBnm@ server developed by Reuter group [164], a service that performs an analysis of the first

Normal Mode with an efficient approach [165, 166]. It uses MMTK library [167] and an appropriate force field [168]. The three first lower modes, *i.e.*, modes 7 to 9, showed clearly that *ECD1,* is involved in large domain motions with the largest contributions to the amplitude of motions (see Figure 8) with the largest contributions. The directions of motions indicated that *ECD1* tends to come closer to the other *ECDs*. Interestingly, this approach highlights the distinction in three regions already observed with a first structured zone, a transition region and a last structured zone. The median region seems to play a role of hinge between the two extremities of *ECD1* (see Figure 8b). However, these results even if quite interesting has to be considered with caution. Indeed, normal mode analysis strongly depends on the 3D structure. So, further analyses have to be performed on alternative models to confirm the ECD1 motion tendencies.

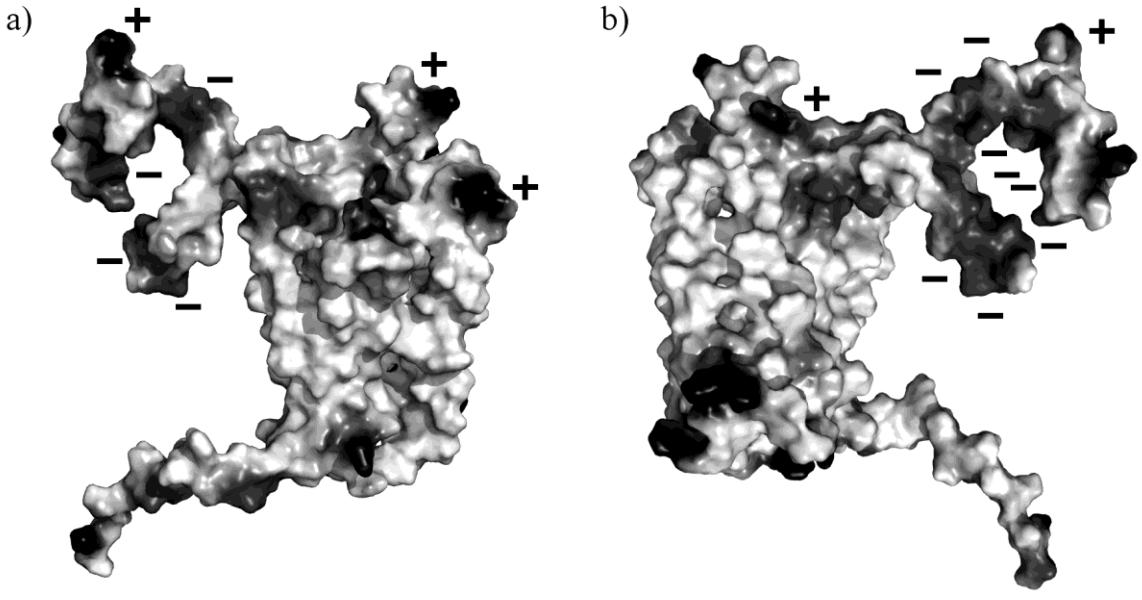

Figure 9. Visualization of electrostatics potential on the surface of DARC 'open' model. Visualization done with PyMol software [160].

Electrostatic potentials of DARC model were calculated using the finite-difference

Poisson-Boltzmann (FDPB) method [169]. Figure 9 present the electrostatic potential mapped onto the molecular surface of the DARC model. Two distinct zones can be observed for the potential interaction zones of CXCL8. The first one is highly negative, the second one is highly positive. The residues in the core of the negative zone are the F22-E23 (corresponding to important epitope Fy6). Interestingly, electrostatics analysis performed for the CXCL8 dimer and monomer showed a complementary with this distribution: two major regions in terms of electrostatic potential can be delineated, one positive and the remainder negative, encompassing the loop 40s. Electrostatic potential computed for CCL5 (RANTES) showed less contrasted results; moreover the charge repartition was very different from CXCL8.

These analyses corroborate other studies that highlighted the importance of electrostatics in the binding of chemokines [170, 171]. For instance, the region of interaction of CXCR-1 which has been partially co-crystallized is positively charged, as the ECD1 N-terminus, and interacts with the loop 40s, negatively charged [172]. So, these results are coherent with all the available data.

Overall, these results show that the models we proposed provide some interesting clues for locating the binding site for natural ligands and the forces that could govern the interaction. However, a precise location, *i.e.* the ensemble of residues that participate to the interaction, necessitates more sophisticated methods such as docking methods. Docking algorithms allow scoring of each position of the ligand rolling onto the entire surface of the receptor. It is necessary to remind that, in case of membrane proteins, the membrane hides numerous sites of the protein. So, a systematic exploration of the receptor surface will waste a lot of time in non relevant locations. Most docking tools first consider the partners as rigid bodies and eventually introduce flexibility at the end of the procedure and generally only for some selected conformations of the complex. Eventually, the ligand (the smallest partner) can be considered as flexible. Flexibility is often limited to side chains repositioning. In some cases,

the polypeptide backbone is optimized by few steps of energy minimization. Thus, taking profit from the ICM software (reference) that allows introducing flexibility from the very first stages, we develop an original strategy to predict binding sites and complex conformations of DARC with natural ligands, *i.e.* CXCL8.

The docking approach

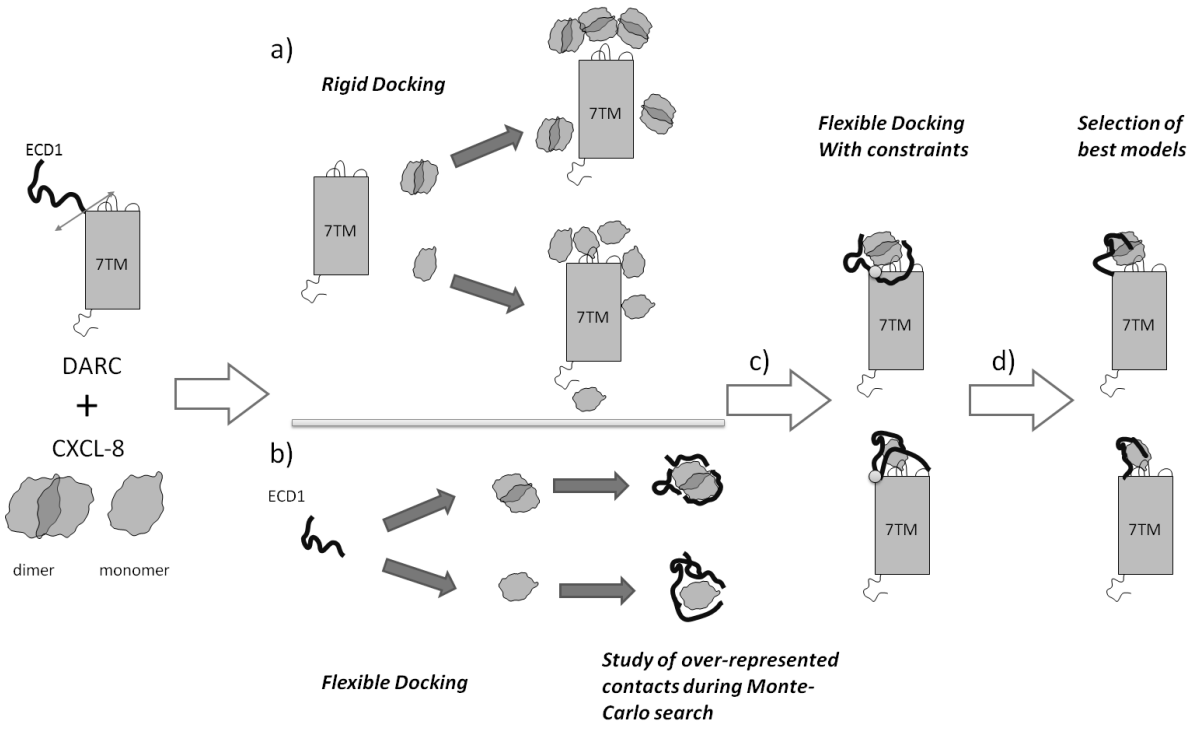

Figure 10. A divide-and-conquer approach. The *ECD1* (N terminus of DARC) and the Transmembrane Domain (TMD) are considered separately. (a) A rigid docking is performed on the TMD while (b) a flexible docking is done for *ECD1*. Then (c) a flexible docking is done to combine both results. (d) the most interesting results are analyzed.

As shown in the previous sections, DARC is a highly complex protein. To take into account the high flexibility of *ECD1*, we have designed a dedicated docking approach that combines rigid and flexible docking. First, a rigid-body docking was performed between CXCL8 and DARC without *ECD1*, *i.e.*, Transmembrane Domain (TMD) encompassing the seven transmembrane helices, the extracellular connecting loops *ECD2*, *ECD3*, *ECD4* and all the *ICDs* intracellular connecting loops (see Figure 10a), even if these last ones do not

participate in the binding. The presence of all the *ECDs* is very important. Indeed, the shape of the receptor molecular surface is dependent on their presence.

In parallel, a flexible docking was performed considering *ECD1* domain as flexible rolling onto the surface of CXCL8 structures, maintained rigid (see Figure 10b). As no clear experimental evidence asserts if the interaction concerns mono- or dimeric form of CXCL8, both hypotheses were tested for rigid and flexible docking. The best results of each approach were then combined and a flexible docking is done where only *ECD1* is flexible (see Figure 10c). Then the best results are selected and analyzed (see Figure 10d).

Rigid docking

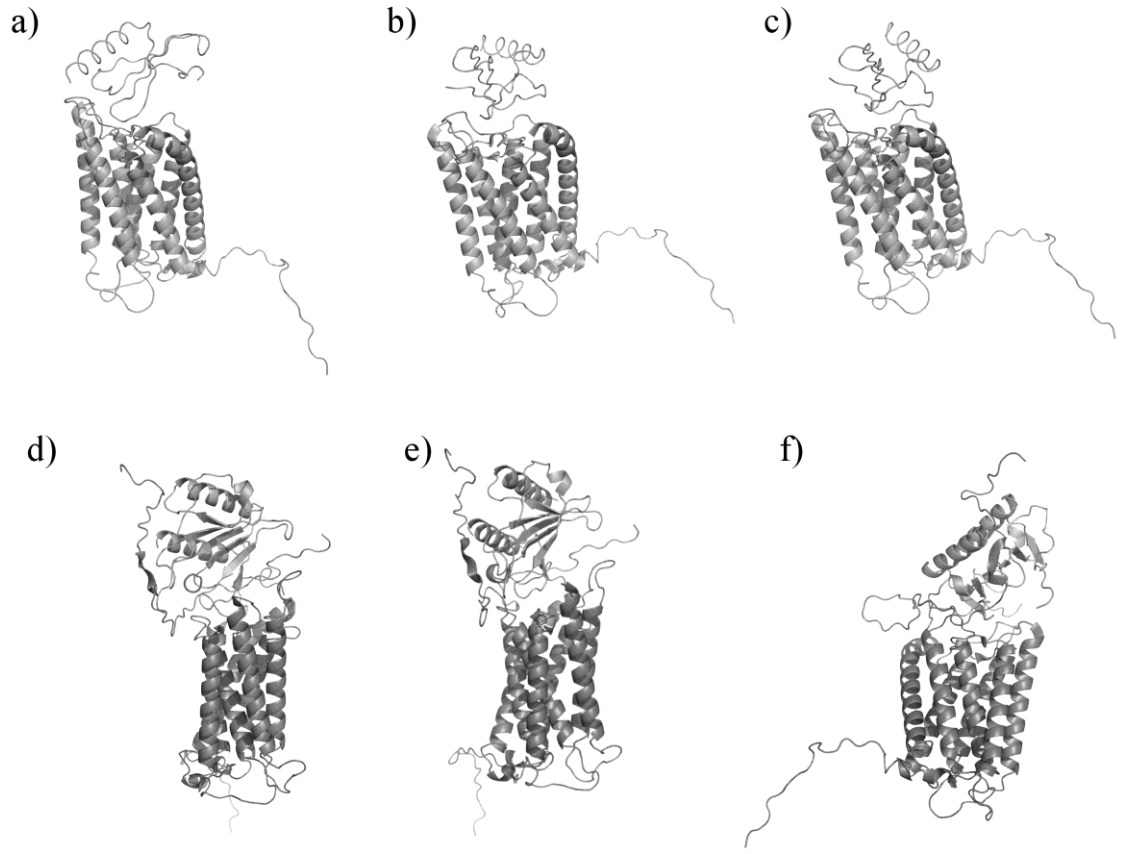

Figure 11. Rigid docking examples. (a – c) Three different examples of rigid docking results between TMD and CXCL8 (monomer form). (d - f) Three different views of a complex obtained by rigid body between TMD and CXCL8 (dimer form).

To screen the different potential interaction conformations between TMD and CXCL8, rigid docking simulations were performed. Two different tools, namely ClusPro [173, 174] and ICM [175, 176], were used. As numerous solutions are generated by each method, clusters of equivalent solutions have been done. Then different filters were applied to select the best solution. A first filter consists in selecting clusters where the ligand is located on the extracellular side, *i.e*. discarding conformations where ligand was in contact with *ICDs* or with the TMD. Indeed, it was not possible to introduce constraints to avoid *a priori* these positioning. Then, we applied experimental filters, *i.e*. the residues 51, 276, 129, 195 … are known to be potentially involved directly in the interaction between DARC and CXCL8 (see Figures 11a to 11d for some examples). One position of CXCL8 was in good agreement with the available biological data (see Figure 12). Moreover, in this conformation CXCL8 is able to bind glycosaminoglycans (GAG). GAGs - chemokine interaction play a key role in the aggregation of chemokines, in the gradient-generating release of the chemokines in the circulation, and their presentation to the receptors [177-179]. The modulation of such interactions may represent a therapeutic approach in inflammatory disease [180]. Around 400 different conformations were analyzed during this step. Results for CXCL8 monomer and dimer were both compatible.

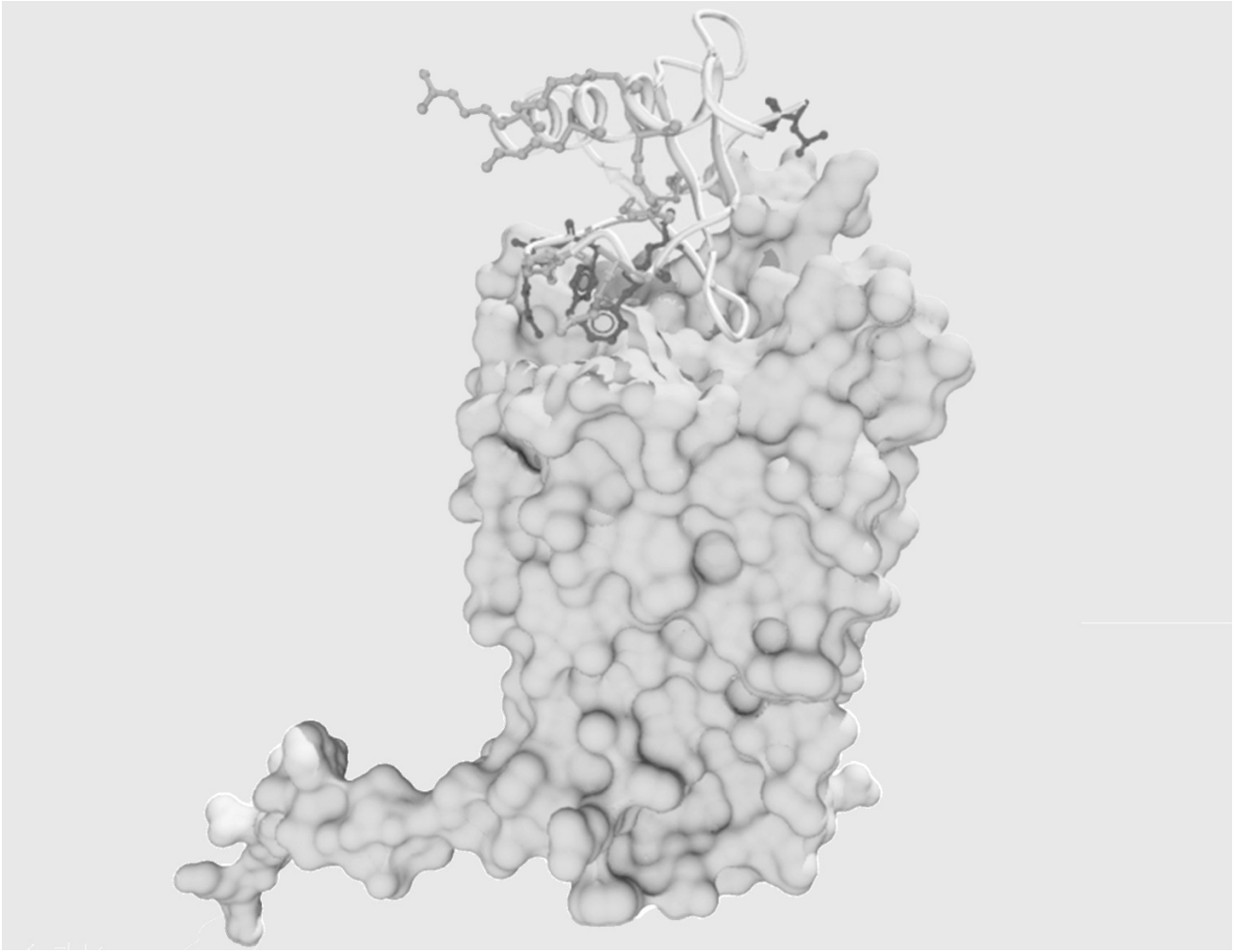

Figure 12. Selected monomer CXCL8 – DARC TMD docking. Some important residues are highlighted.

Flexible docking

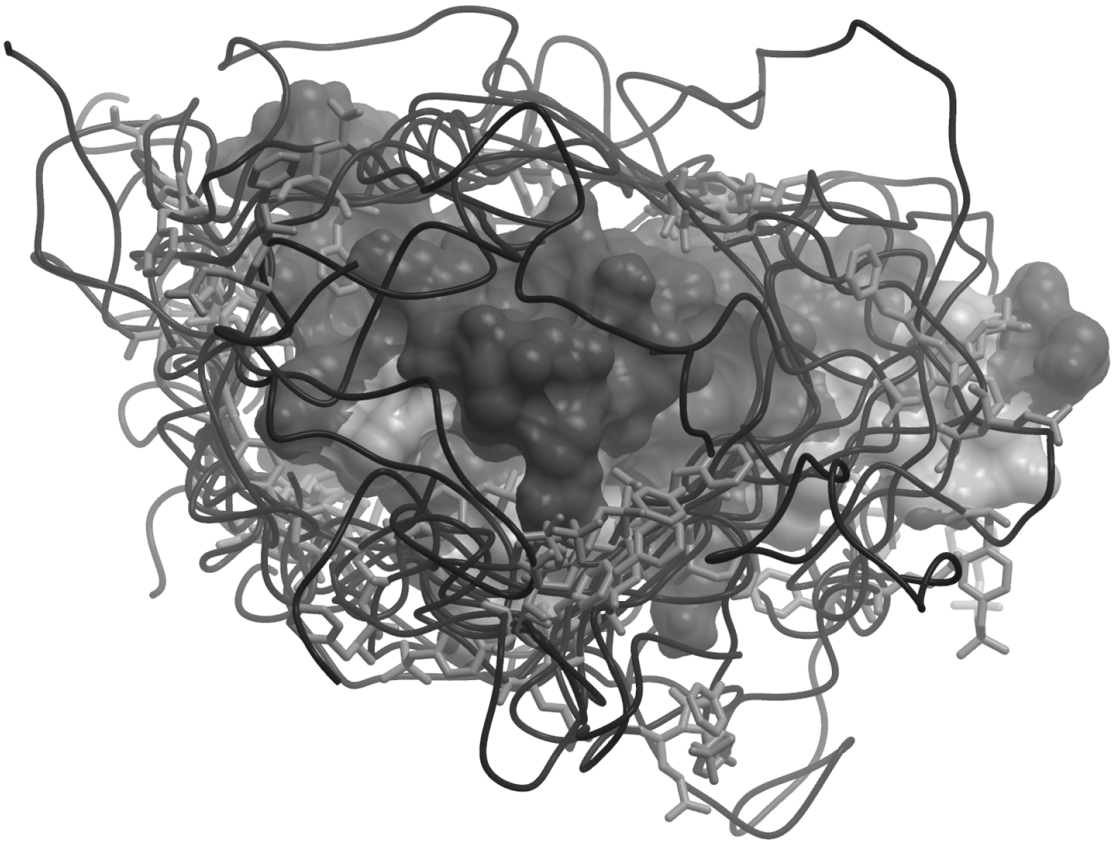

Figure 13. Flexible docking. Some snapshots of the flexible search of *ECD1* around the rigid structure of CXCL8 monomer are shown.

In parallel to the rigid docking, a docking was done where *ECD1* was allowed to adapt its conformation when interacting with CXCL8. This flexible docking was done using the biased Monte-Carlo (MC) search of ICM [175, 176]based on a. *ECD1* is firstly placed by a preliminary rigid body docking. Then, a random sampling is done in the ϕ-ψ space, *i.e.*, one couple ϕ-ψ is chosen in the favored Ramachandran regions and affected to one residue. The corresponding energy is computed, and then the Metropolis criterion is applied to keep or discard the conformation so generated. One million of random has been performed and 250.000 MC steps. Privileged conformations were clustered, leading to 50 clusters (see Figure 13). Three clusters were in accordance with the biological data, *i.e.* residues of CXCL8 and of DARC supposed to be implicated in the interaction. The contacts observed between *ECD1*

and CXCL8, were analyzed using Protein Blocks approach [138, 139].

Complete assembly

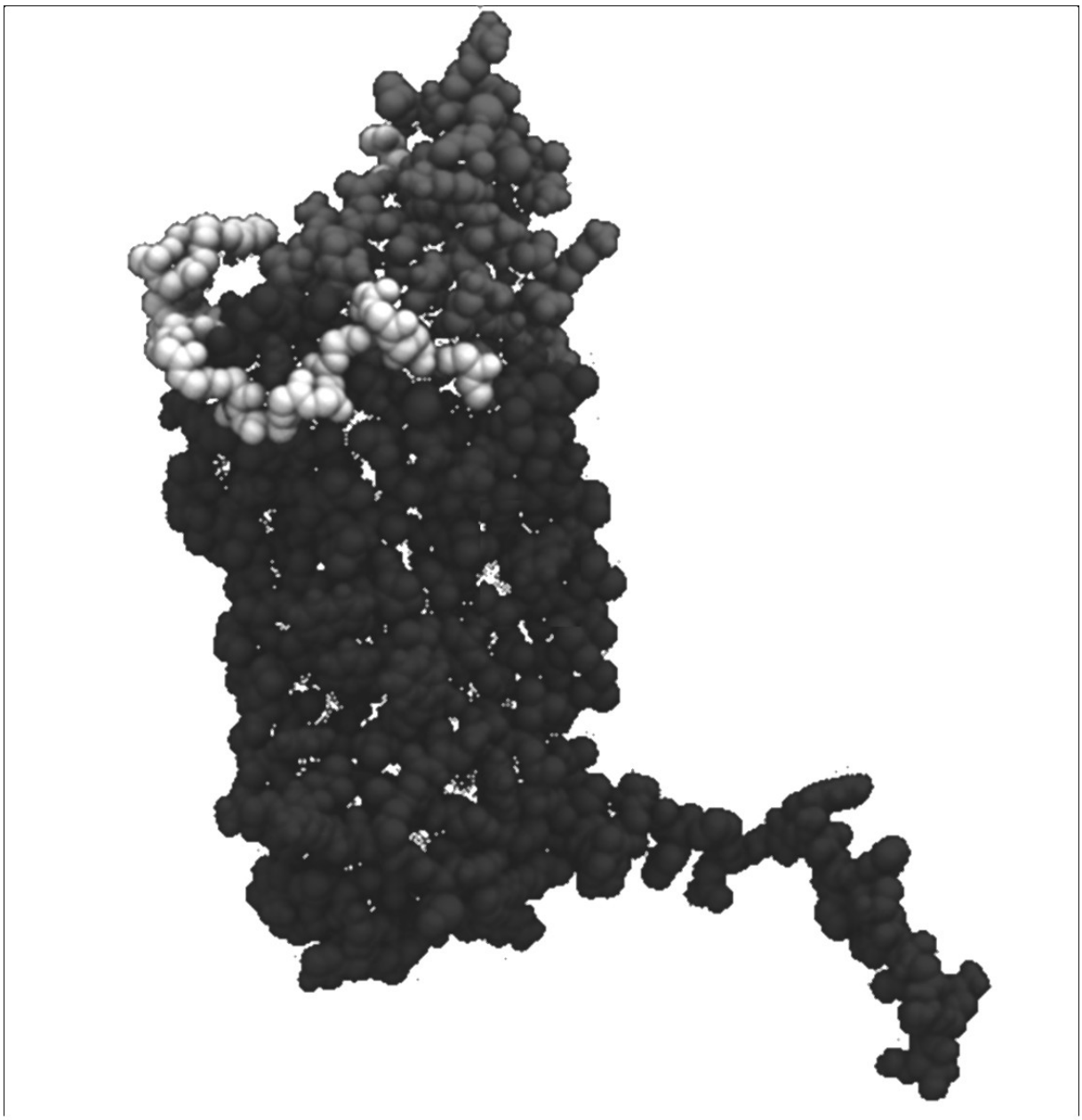

Figure 14. A good conformation. ECD1, CXCL8 and TD are shown with different grey tones.

Spatial position of CXCL8 obtained from the rigid docking was combined with the results of flexible docking. The main task consisted in selecting *ECD1* conformations where the peptide bond between the last residue of *ECD1* and the first residue of TMD can be easily

made. The closure of the bond was obtained by a minimization with constraints. A new flexible docking (only *ECD1* was flexible) was done to take into account the whole system. This simulation generates ~120.000 conformations which were grouped and filtered according to the biological data. For examples, 62 clusters of distinct conformations were selected for DARC Tyrosine 30 with CXCL8. Three conformations were associated to the lowest energy. Figure 14 shows the most representative conformation (the most frequently observed during the simulation) and that presenting the best compromise between favorable energy versus number of occurrences (see Figure 14).

Overall, even if the results are preliminary, the strategy described here yields interesting information that complement the biological data available.

Future works

The selected models need to be deeply and carefully analyzed and especially at the level of the amino acids involved at the interface between DARC and CXCL8. Concerning the DBP, the structure is incomplete but available [181]. This structure brings some hints [182, 183] to understand the invasion mechanism. A preliminary study performed for DARC-DBP complex, using a similar approach to that described above give remarkable results highly compatible with the main interaction hypothesis [183]. Models need to be refined using classical molecular modeling. Moreover, we have identified different pockets interesting for the interaction; further deeper analyses would help to understand their roles and their importance in the binding. Using these different results, it will be possible to propose designed mutants that can be tested first *in silico*.

Acknowledgments

This work was supported by National Institute for Health and Medical Care (INSERM), University Paris 7 – Denis Diderot and National Institute for Blood Transfusion (INTS). LA benefits from a grant of the Région Ile-De-France. AdB and OB acknowledge Polonium grant 14287VM "Refinement of the studies on specificity of monoclonal antibody 2C3 directed against Duffy/DARC protein".

Figure Legends

Figure 1. The three steps. (1) by comparative modeling, structural models are building; (2) docking approach is performed with the natural ligand structure, *i.e.*, CXC-L8 and (3) with the Duffy Binding Protein (DBP).

Figure 2. Principles of homology modeling for a transmembrane protein.

Figure 3. Phylogenetic analysis of DARC. Different sequences of DARC were aligned using CLUSTALW [91].The dendogram is plotted with Njplot software [92]. Bootstrap values computed by CLUSTALW are shown at the different nodes. Sequences were extracted from UniProt/UniParc ID [94]: the major 336 aa [93] human (DUFF_HUMAN, Q16570 [95]) was used, gorilla (*Gorilla gorilla*), AF311914; marmoset (*Callithrix jacchus*), AF311915; tamarin (*Saguinus oedipus*), AF311916; night monkey (*Aotus trivirgatus*), AF311917; squirrel monkey (*Samiri boliviensis*), AF311918; brown capuchin (*Cebus apella*), AF311919; chimpanzee (*Pan troglodytes*), AF311920; rhesus monkey (*Macaca mulatta*), AF311921; baboon (*Papio papio*), AF303532; gibbon (*Hylobates lar*) AF303533 (all the last sequences are from [71]).

Figure 4. Prediction of transmembrane segments of DARC. Are shown the predictions done using HMMTOP [96, 97], TMHMM 2.0 [98], TMAP [113], MEMSAT [117], SOSUI [107], TMpred [104], TSEG [114], TM-FINDER [115], Pred-TMR 2.0 [111], SPLIT [109] and DAS [99]. The final selection corresponds to the positions finally obtained after multiple tests and evaluations. The plot was done using R software [121].

Figure 5. Building structural models of DARC. (a) Prediction of transmembrane helices. (b) Alignment of helical regions with corresponding regions of rhodopsin structure. (c) Potential structural templates for .*ECD1* and *ICD4* prediction of (d) Addition of these results to the complete alignment for comparative modeling. (e) Structural model generation and refinement of these models. (f) Accessibility computation of amino acids and known to be exposed. (g) In regards to the results, the alignment is modified. (h) At last, some models are selected.

Figure 6. Encoding of the protein structures (3D) in terms of Protein Blocks.

Figure 7. The two DARC selected models. Two views of (a-b) open form structural model and (c-d) closed structural model. Visualization done with PyMol software [160].

Figure 8. Normal Mode Analysis of DARC 'open' model. (a) (y-axis) Modes 7 to 9 for (x-axis) the whole sequence. (b) Visualization of the modes 7 to 9 on the 'open' model. The cross highlights the hinge region of *ECD1*. Visualization done with VMD software [162, 163].

Figure 9. Visualization of electrostatics potential on the surface of DARC 'open' model. Visualization done with PyMol software [160].

Figure 10. A divide-and-conquer approach. The *ECD1* (N terminus of DARC) and the Transmembrane Domain (TMD) are considered separately. (a) A rigid docking is performed on the TMD while (b) a flexible docking is done for *ECD1*. Then (c) a flexible docking is done to combine both results. (d) the most interesting results are analyzed.

Figure 11. Rigid docking examples. (a – c) Three different examples of rigid docking results between TMD and CXCL8 (monomer form). (d - f) Three different views of a complex obtained by rigid body between TMD and CXCL8 (dimer form).

Figure 12. Selected monomer CXCL8 – DARC TMD docking. Some important residues are

highlighted.

Figure 13. Flexible docking. Some snapshots of the flexible search of *ECD1* around the rigid structure of CXCL8 monomer are shown.

Figure 14. A good conformation. ECD1, CXCL8 and TD are shown with different grey tones.

References

[1] Cutbush M., Mollison P.L., The Duffy blood group system, Heredity 4 (1950) 383-389.

[2] Ikin E.W., Mourant A.E., Pettenkofer H.J., Blumenthal G., Discovery of the expected haemagglutinin, anti-Fyb, Nature 168 (1951) 1077-1078.

[3] Albrey J.A., Vincent E.E., Hutchinson J., Marsh W.L., Allen F.H., Jr., Gavin J., Sanger R., A new antibody, anti-Fy3, in the Duffy blood group system, Vox Sang 20 (1971) 29-35.

[4] Chown B., Lewis M., Kaita H., The Duffy Blood Group System in Caucasians: Evidence for a New Allele, Am J Hum Genet 17 (1965) 384-389.

[5] Compton A., Haber J.M., The duffy blood group system in transfusion reactions: a reviw of the literature and report of four cases, Blood 15 (1960) 186-191.

[6] Gawrzewski W., Kalczew J., Studies on the Duffy blood group system (FYa) in the population of Cracow (Poland), Acta Med Pol 6 (1965) 255-256.

[7] Lewis M., Kaita H., Chown B., The Duffy blood group system in Caucasians. A further population sample, Vox Sang 23 (1972) 523-527.

[8] Marsh W.L., Present status of the Duffy blood group system, CRC Crit Rev Clin Lab Sci 5 (1975) 387-412.

[9] Marsh W.L., Ehrich C.C., The Duffy blood group system: a review of recent developments, Infusionsther Klin Ernahr 2 (1975) 280-289.

[10] Mohr J., Note on the inheritance of the Duffy blood-group system and its possible interaction with the Rhesus groups, Ann Eugen 18 (1954) 318-324.

[11] Oberdorfer C.E., Kahn B., Moore V., Zelenski K., Oyen R., Marsh W.L., A second example of anti-Fy3 in the Duffy blood group system, Transfusion 14 (1974) 608-611.

[12] Beattie K., A review: the Duffy blood group system, Immunohematology 5 (1989) 45-54.

[13] Donahue R.P., Bias W.B., Renwick J.H., McKusick V.A., Probable assignment of the Duffy blood group locus to chromosome 1 in man, Proc Natl Acad Sci U S A 61 (1968) 949-955.

[14] Miller L.H., Mason S.J., Dvorak J.A., McGinniss M.H., Rothman I.K., Erythrocyte receptors for (Plasmodium knowlesi) malaria: Duffy blood group determinants, Science 189 (1975) 561-563.

[15] Miller L.H., Mason S.J., Clyde D.F., McGinniss M.H., The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy, N Engl J Med 295 (1976) 302-304.

[16] Tournamille C., Colin Y., Cartron J.P., Le Van Kim C., Disruption of a GATA motif

in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals, Nat Genet 10 (1995) 224-228.

[17] Carlton J.M., Adams J.H., Silva J.C., Bidwell S.L., Lorenzi H., Caler E., Crabtree J., Angiuoli S.V., Merino E.F., Amedeo P., Cheng Q., Coulson R.M., Crabb B.S., Del Portillo H.A., Essien K., Feldblyum T.V., Fernandez-Becerra C., Gilson P.R., Gueye A.H., Guo X., Kang'a S., Kooij T.W., Korsinczky M., Meyer E.V., Nene V., Paulsen I., White O., Ralph S.A., Ren Q., Sargeant T.J., Salzberg S.L., Stoeckert C.J., Sullivan S.A., Yamamoto M.M., Hoffman S.L., Wortman J.R., Gardner M.J., Galinski M.R., Barnwell J.W., Fraser-Liggett C.M., Comparative genomics of the neglected human malaria parasite Plasmodium vivax, Nature 455 (2008) 757-763.

[18] Guerra C.A., Snow R.W., Hay S.I., Mapping the global extent of malaria in 2005, Trends Parasitol 22 (2006) 353-358.

[19] Mendis K., Sina B.J., Marchesini P., Carter R., The neglected burden of Plasmodium vivax malaria, Am J Trop Med Hyg 64 (2001) 97-106.

[20] Sattabongkot J., Tsuboi T., Zollner G.E., Sirichaisinthop J., Cui L., Plasmodium vivax transmission: chances for control?, Trends Parasitol 20 (2004) 192-198.

[21] Talib V.H., Kiran P.C., Talib N.J., Choudhury M., Chloroquine-resistant Plasmodium vivax malaria in infancy and childhood, Indian J Pediatr 46 (1979) 158-162.

[22] Whitby M., Wood G., Veenendaal J.R., Rieckmann K., Chloroquine-resistant Plasmodium vivax, Lancet 2 (1989) 1395.

[23] de Santana Filho F.S., Arcanjo A.R., Chehuan Y.M., Costa M.R., Martinez-Espinosa F.E., Vieira J.L., Barbosa M.G., Alecrim W.D., Alecrim M.G., Chloroquine-resistant Plasmodium vivax, Brazilian Amazon, Emerg Infect Dis 13 (2007) 1125-1126.

[24] Dua V.K., Kar P.K., Sharma V.P., Chloroquine resistant Plasmodium vivax malaria in India, Trop Med Int Health 1 (1996) 816-819.

[25] Teka H., Petros B., Yamuah L., Tesfaye G., Elhassan I., Muchohi S., Kokwaro G., Aseffa A., Engers H., Chloroquine-resistant Plasmodium vivax malaria in Debre Zeit, Ethiopia, Malar J 7 (2008) 220.

[26] Kurcer M.A., Simsek Z., Kurcer Z., The decreasing efficacy of chloroquine in the treatment of Plasmodium vivax malaria, in Sanliurfa, south-eastern Turkey, Ann Trop Med Parasitol 100 (2006) 109-113.

[27] Kurcer M.A., Simsek Z., Zeyrek F.Y., Atay S., Celik H., Kat I., Topluoglu S., Efficacy of chloroquine in the treatment of Plasmodium vivax malaria in Turkey, Ann Trop Med Parasitol 98 (2004) 447-451.

[28] Sumawinata I.W., Bernadeta, Leksana B., Sutamihardja A., Purnomo, Subianto B., Sekartuti, Fryauff D.J., Baird J.K., Very high risk of therapeutic failure with chloroquine for uncomplicated Plasmodium falciparum and P. vivax malaria in Indonesian Papua, Am J Trop Med Hyg 68 (2003) 416-420.

[29] Girod R., Gaborit P., Carinci R., Issaly J., Fouque F., Anopheles darlingi bionomics and transmission of Plasmodium falciparum, Plasmodium vivax and Plasmodium malariae in Amerindian villages of the Upper-Maroni Amazonian forest, French Guiana, Mem Inst Oswaldo Cruz 103 (2008) 702-710.

[30] Beeson J.G., Crabb B.S., Towards a vaccine against Plasmodium vivax malaria, PLoS Med 4 (2007) e350.

[31] Grimberg B.T., Udomsangpetch R., Xainli J., McHenry A., Panichakul T., Sattabongkot J., Cui L., Bockarie M., Chitnis C., Adams J., Zimmerman P.A., King C.L., Plasmodium vivax invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein, PLoS Med 4 (2007) e337.

[32] Moreno A., Caro-Aguilar I., Yazdani S.S., Shakri A.R., Lapp S., Strobert E., McClure H., Chitnis C.E., Galinski M.R., Preclinical assessment of the receptor-binding domain

of Plasmodium vivax Duffy-binding protein as a vaccine candidate in rhesus macaques, Vaccine 26 (2008) 4338-4344.

[33] King C.L., Michon P., Shakri A.R., Marcotty A., Stanisic D., Zimmerman P.A., Cole-Tobian J.L., Mueller I., Chitnis C.E., Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage Plasmodium vivax infection, Proc Natl Acad Sci U S A 105 (2008) 8363-8368.

[34] Ryan J.R., Stoute J.A., Amon J., Dunton R.F., Mtalib R., Koros J., Owour B., Luckhart S., Wirtz R.A., Barnwell J.W., Rosenberg R., Evidence for transmission of Plasmodium vivax among a duffy antigen negative population in Western Kenya, Am J Trop Med Hyg 75 (2006) 575-581.

[35] Rosenberg R., Plasmodium vivax in Africa: hidden in plain sight?, Trends Parasitol 23 (2007) 193-196.

[36] Culleton R.L., Mita T., Ndounga M., Unger H., Cravo P.V., Paganotti G.M., Takahashi N., Kaneko A., Eto H., Tinto H., Karema C., D'Alessandro U., do Rosario V., Kobayakawa T., Ntoumi F., Carter R., Tanabe K., Failure to detect Plasmodium vivax in West and Central Africa by PCR species typing, Malar J 7 (2008) 174.

[37] Cavasini C.E., de Mattos L.C., Couto A.A., Couto V.S., Gollino Y., Moretti L.J., Bonini-Domingos C.R., Rossit A.R., Castilho L., Machado R.L., Duffy blood group gene polymorphisms among malaria vivax patients in four areas of the Brazilian Amazon region, Malar J 6 (2007) 167.

[38] Darbonne W.C., Rice G.C., Mohler M.A., Apple T., Hebert C.A., Valente A.J., Baker J.B., Red blood cells are a sink for interleukin 8, a leukocyte chemotaxin, J Clin Invest 88 (1991) 1362-1369.

[39] Horuk R., Chitnis C.E., Darbonne W.C., Colby T.J., Rybicki A., Hadley T.J., Miller L.H., A receptor for the malarial parasite Plasmodium vivax: the erythrocyte chemokine receptor, Science 261 (1993) 1182-1184.

[40] Murdoch C., Finn A., Chemokine receptors and their role in inflammation and infectious diseases, Blood 95 (2000) 3032-3043.

[41] Neote K., Mak J.Y., Kolakowski L.F., Jr., Schall T.J., Functional and biochemical analysis of the cloned Duffy antigen: identity with the red blood cell chemokine receptor, Blood 84 (1994) 44-52.

[42] Neote K., DiGregorio D., Mak J.Y., Horuk R., Schall T.J., Molecular cloning, functional expression, and signaling characteristics of a C-C chemokine receptor, Cell 72 (1993) 415-425.

[43] Chaudhuri A., Zbrzezna V., Polyakova J., Pogo A.O., Hesselgesser J., Horuk R., Expression of the Duffy antigen in K562 cells. Evidence that it is the human erythrocyte chemokine receptor, J Biol Chem 269 (1994) 7835-7838.

[44] Szabo M.C., Soo K.S., Zlotnik A., Schall T.J., Chemokine class differences in binding to the Duffy antigen-erythrocyte chemokine receptor, J Biol Chem 270 (1995) 25348-25351.

[45] Comerford I., Nibbs R.J., Post-translational control of chemokines: a role for decoy receptors?, Immunol Lett 96 (2005) 163-174.

[46] Hadley T.J., Peiper S.C., From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen, Blood 89 (1997) 3077-3091.

[47] Wasniowska K., Hadley T.J., The Duffy blood group antigen: an update, Transfus Med Rev 8 (1994) 281-288.

[48] Graham G.J., D6 and the atypical chemokine receptor family: novel regulators of immune and inflammatory processes, Eur J Immunol 39 (2009) 342-351.

[49] Pruenster M., Rot A., Throwing light on DARC, Biochem Soc Trans 34 (2006) 1005-1008.

[50] Chakera A., Seeber R.M., John A.E., Eidne K.A., Greaves D.R., The duffy antigen/receptor for chemokines exists in an oligomeric form in living cells and functionally antagonizes CCR5 signaling through hetero-oligomerization, Mol Pharmacol 73 (2008) 1362-1370.

[51] Pruenster M., Mudde L., Bombosi P., Dimitrova S., Zsak M., Middleton J., Richmond A., Graham G.J., Segerer S., Nibbs R.J., Rot A., The Duffy antigen receptor for chemokines transports chemokines and supports their promigratory activity, Nat Immunol 10 (2009) 101-108.

[52] Bruhl H., Vielhauer V., Weiss M., Mack M., Schlondorff D., Segerer S., Expression of DARC, CXCR3 and CCR5 in giant cell arteritis, Rheumatology (Oxford) 44 (2005) 309-313.

[53] Gardner L., Wilson C., Patterson A.M., Bresnihan B., FitzGerald O., Stone M.A., Ashton B.A., Middleton J., Temporal expression pattern of Duffy antigen in rheumatoid arthritis: up-regulation in early disease, Arthritis Rheum 54 (2006) 2022-2026.

[54] Lappin D.W., Brady H.R., Light from DARCness, Kidney Int 58 (2000) 1816-1817.

[55] Lee J.S., Frevert C.W., Wurfel M.M., Peiper S.C., Wong V.A., Ballman K.K., Ruzinski J.T., Rhim J.S., Martin T.R., Goodman R.B., Duffy antigen facilitates movement of chemokine across the endothelium in vitro and promotes neutrophil transmigration in vitro and in vivo, J Immunol 170 (2003) 5244-5251.

[56] Liu X.H., Hadley T.J., Xu L., Peiper S.C., Ray P.E., Up-regulation of Duffy antigen receptor expression in children with renal disease, Kidney Int 55 (1999) 1491-1500.

[57] Patterson A.M., Siddall H., Chamberlain G., Gardner L., Middleton J., Expression of the duffy antigen/receptor for chemokines (DARC) by the inflamed synovial endothelium, J Pathol 197 (2002) 108-116.

[58] Segerer S., Regele H., Mac K.M., Kain R., Cartron J.P., Colin Y., Kerjaschki D., Schlondorff D., The Duffy antigen receptor for chemokines is up-regulated during acute renal transplant rejection and crescentic glomerulonephritis, Kidney Int 58 (2000) 1546-1556.

[59] Vandercappellen J., Van Damme J., Struyf S., The role of CXC chemokines and their receptors in cancer, Cancer Lett 267 (2008) 226-244.

[60] Lentsch A.B., The Duffy antigen/receptor for chemokines (DARC) and prostate cancer. A role as clear as black and white?, FASEB J 16 (2002) 1093-1095.

[61] Shen H., Schuster R., Stringer K.F., Waltz S.E., Lentsch A.B., The Duffy antigen/receptor for chemokines (DARC) regulates prostate tumor growth, FASEB J 20 (2006) 59-64.

[62] Addison C.L., Belperio J.A., Burdick M.D., Strieter R.M., Overexpression of the duffy antigen receptor for chemokines (DARC) by NSCLC tumor cells results in increased tumor necrosis, BMC Cancer 4 (2004) 28.

[63] Wang J., Ou Z.L., Hou Y.F., Luo J.M., Shen Z.Z., Ding J., Shao Z.M., Enhanced expression of Duffy antigen receptor for chemokines by breast cancer cells attenuates growth and metastasis potential, Oncogene 25 (2006) 7201-7211.

[64] Vergara C., Tsai Y.J., Grant A.V., Rafaels N., Gao L., Hand T., Stockton M., Campbell M., Mercado D., Faruque M., Dunston G., Beaty T.H., Oliveira R.R., Ponte E.V., Cruz A.A., Carvalho E., Araujo M.I., Watson H., Schleimer R.P., Caraballo L., Nickel R.G., Mathias R.A., Barnes K.C., Gene encoding Duffy antigen/receptor for chemokines is associated with asthma and IgE in three populations, Am J Respir Crit Care Med 178 (2008) 1017-1022.

[65] He W., Neil S., Kulkarni H., Wright E., Agan B.K., Marconi V.C., Dolan M.J., Weiss R.A., Ahuja S.K., 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 4 (2008) 52-62.
[66] Walton R.T., Rowland-Jones S.L., HIV and chemokine binding to red blood cells--DARC matters, Cell Host Microbe 4 (2008) 3-5.
[67] Editorial Team, Study links genetic polymorphism to susceptibility to HIV infection, Euro Surveill 13 (2008).
[68] Smith M., The year in human and medical genetics. Highlights of 2007-2008, Ann N Y Acad Sci 1151 (2009) 1-21.
[69] Le Van Kim C., Tournamille C., Kroviarski Y., Cartron J.P., 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 90 (1997) 2851-2853.
[70] Tournamille C., [Molecular basis and structure-activity relationships of the Duffy blood group antigens: chemokine and Plasmodium vivax receptors], Transfus Clin Biol 7 (2000) 497-509.
[71] Tournamille C., Blancher A., Le Van Kim C., Gane P., Apoil P.A., Nakamoto W., Cartron J.P., Colin Y., Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines, Immunogenetics 55 (2004) 682-694.
[72] Tournamille C., Filipe A., Wasniowska K., Gane P., Lisowska E., Cartron J.P., Colin Y., Le Van Kim C., Structure-function analysis of the extracellular domains of the Duffy antigen/receptor for chemokines: characterization of antibody and chemokine binding sites, Br J Haematol 122 (2003) 1014-1023.
[73] Tournamille C., Le Van Kim C., Gane P., Blanchard D., Proudfoot A.E., Cartron J.P., Colin Y., 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 272 (1997) 16274-16280.
[74] Tournamille C., Le Van Kim C., Gane P., Le Pennec P.Y., Roubinet F., Babinet J., Cartron J.P., Colin Y., Arg89Cys substitution results in very low membrane expression of the Duffy antigen/receptor for chemokines in Fy(x) individuals, Blood 92 (1998) 2147-2156.
[75] Wasniowska K., Lisowska E., Halverson G.R., Chaudhuri A., Reid M.E., The Fya, Fy6 and Fy3 epitopes of the Duffy blood group system recognized by new monoclonal antibodies: identification of a linear Fy3 epitope, Br J Haematol 124 (2004) 118-122.
[76] 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 356 (2007) 816-821.
[77] Wasniowska K., Blanchard D., Janvier D., Wang Z.X., Peiper S.C., Hadley T.J., Lisowska E., 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 33 (1996) 917-923.
[78] Hadley T.J., Lu Z.H., Wasniowska K., Martin A.W., Peiper S.C., Hesselgesser J., Horuk R., 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 94 (1994) 985-991.
[79] Sim B.K., Chitnis C.E., Wasniowska K., Hadley T.J., Miller L.H., Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum, Science 264 (1994) 1941-1944.
[80] Tournamille C., Filipe A., Badaut C., Riottot M.M., Longacre S., Cartron J.P., Le Van Kim C., Colin Y., Fine mapping of the Duffy antigen binding site for the Plasmodium

vivax Duffy-binding protein, Mol Biochem Parasitol 144 (2005) 100-103.
[81] de Brevern A.G., 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 1724 (2005) 288-306.
[82] Singh K., Gittis A.G., Nguyen P., Gowda D.C., Miller L.H., Garboczi D.N., Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A, Nat Struct Mol Biol 15 (2008) 932-938.
[83] Klabunde T., Hessler G., Drug design strategies for targeting G-protein-coupled receptors, Chembiochem 3 (2002) 928-944.
[84] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E., The Protein Data Bank, Nucleic Acids Res 28 (2000) 235-242.
[85] Fleishman S.J., Unger V.M., Ben-Tal N., Transmembrane protein structures without X-rays, Trends Biochem Sci 31 (2006) 106-113.
[86] Radestock S., Weil T., Renner S., Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring, J Chem Inf Model 48 (2008) 1104-1117.
[87] Paterlini M.G., Structure modeling of the chemokine receptor CCR5: implications for ligand binding and selectivity, Biophys J 83 (2002) 3012-3031.
[88] Palczewski K., Kumasaka T., Hori T., Behnke C.A., Motoshima H., Fox B.A., Le Trong I., Teller D.C., Okada T., Stenkamp R.E., Yamamoto M., Miyano M., Crystal structure of rhodopsin: A G protein-coupled receptor, Science 289 (2000) 739-745.
[89] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res 25 (1997) 3389-3402.
[90] Pearson W.R., Lipman D.J., Improved tools for biological sequence comparison, Proc Natl Acad Sci U S A 85 (1988) 2444-2448.
[91] Thompson J.D., Higgins D.G., Gibson T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res 22 (1994) 4673-4680.
[92] Perriere G., Gouy M., WWW-query: an on-line retrieval system for biological sequence banks, Biochimie 78 (1996) 364-369.
[93] Iwamoto S., Li J., Omi T., Ikemoto S., Kajii E., Identification of a novel exon and spliced form of Duffy mRNA that is the predominant transcript in both erythroid and postcapillary venule endothelium, Blood 87 (1996) 378-385.
[94] Apweiler R., Bairoch A., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovan C., Redaschi N., Yeh L.S., UniProt: the Universal Protein knowledgebase, Nucleic Acids Res 32 (2004) D115-119.
[95] Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L., GenBank, Nucleic Acids Res 36 (2008) D25-30.
[96] Tusnady G.E., Simon I., Principles governing amino acid composition of integral membrane proteins: application to topology prediction, J Mol Biol 283 (1998) 489-506.
[97] Tusnady G.E., Simon I., The HMMTOP transmembrane topology prediction server, Bioinformatics 17 (2001) 849-850.
[98] Krogh A., Larsson B., von Heijne G., Sonnhammer E.L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J Mol Biol 305 (2001) 567-580.
[99] 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 10 (1997) 673-676.
[100] von Heijne G., Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule, J Mol Biol 225 (1992) 487-494.
[101] Claros M.G., von Heijne G., TopPred II: an improved software for membrane protein structure predictions, Comput Appl Biosci 10 (1994) 685-686.
[102] Rost B., Casadio R., Fariselli P., Sander C., Transmembrane helices predicted at 95% accuracy, Protein Sci 4 (1995) 521-533.
[103] Rost B., Fariselli P., Casadio R., Topology prediction for helical transmembrane proteins at 86% accuracy, Protein Sci 5 (1996) 1704-1718.
[104] Hofmann K., Stoffel W., TMbase - A database of membrane spanning proteins segments, Biol. Chem. Hoppe-Seyler 374 (1993) 166-169.
[105] Hirokawa T., Boon-Chieng S., Mitaku S., SOSUI: classification and secondary structure prediction system for membrane proteins, Bioinformatics 14 (1998) 378-379.
[106] Mitaku S., Hirokawa T., Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length, Protein Eng 12 (1999) 953-957.
[107] 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 18 (2002) 608-616.
[108] Juretic D., Lee B., Trinajstic N., Williams R.W., Conformational preference functions for predicting helices in membrane proteins, Biopolymers 33 (1993) 255-273.
[109] Juretic D., Zoranic L., Zucic D., Basic charge clusters and predictions of membrane protein topology, J Chem Inf Comput Sci 42 (2002) 620-632.
[110] Pasquier C., Hamodrakas S.J., An hierarchical artificial neural network system for the classification of transmembrane proteins, Protein Eng 12 (1999) 631-634.
[111] Pasquier C., Promponas V.J., Palaios G.A., Hamodrakas J.S., Hamodrakas S.J., A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm, Protein Eng 12 (1999) 381-385.
[112] Persson B., Argos P., Prediction of transmembrane segments in proteins utilising multiple sequence alignments, J Mol Biol 237 (1994) 182-192.
[113] Persson B., Argos P., Topology prediction of membrane proteins, Protein Sci 5 (1996) 363-371.
[114] Kihara D., Shimizu T., Kanehisa M., Prediction of membrane proteins based on classification of transmembrane segments, Protein Eng 11 (1998) 961-970.
[115] Deber C.M., Wang C., Liu L.P., Prior A.S., Agrawal S., Muskat B.L., Cuticchia A.J., TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales, Protein Sci 10 (2001) 212-219.
[116] Zhou H., Zhou Y., Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model-based method, Protein Sci 12 (2003) 1547-1555.
[117] Jones D.T., Taylor W.R., Thornton J.M., A model recognition approach to the prediction of all-helical membrane protein structure and topology, Biochemistry 33 (1994) 3038-3049.
[118] Jones D.T., Do transmembrane protein superfolds exist?, FEBS Lett 423 (1998) 281-285.
[119] Viklund H., Elofsson A., Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information,

Protein Sci 13 (2004) 1908-1917.
[120] Shen H., Chou J.J., MemBrain: improving the accuracy of predicting transmembrane helices, PLoS ONE 3 (2008) e2399.
[121] Ihaka R., Gentleman R., R: a language for data analysis and graphics, J Comput Graph Stat 5 (1996) 299-314.
[122] Wasniowska K., Petit-LeRoux Y., Tournamille C., Le van Kim C., Cartron J.P., Colin Y., Lisowska E., Blanchard D., Structural characterization of the epitope recognized by the new anti-Fy6 monoclonal antibody NaM 185-2C3, Transfus Med 12 (2002) 205-211.
[123] Kabsch W., Sander C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577-2637.
[124] Sali A., Blundell T.L., 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 212 (1990) 403-428.
[125] Sali A., Blundell T.L., Comparative protein modelling by satisfaction of spatial restraints, J Mol Biol 234 (1993) 779-815.
[126] Marti-Renom M.A., Stuart A.C., Fiser A., Sanchez R., Melo F., Sali A., Comparative protein structure modeling of genes and genomes, Annu Rev Biophys Biomol Struct 29 (2000) 291-325.
[127] Canutescu A.A., Shelenkov A.A., Dunbrack R.L., Jr., A graph-theory algorithm for rapid protein side-chain prediction, Protein Sci 12 (2003) 2001-2014.
[128] Kutzner C., van der Spoel D., Fechner M., Lindahl E., Schmitt U.W., de Groot B.L., Grubmuller H., Speeding up parallel GROMACS on high-latency networks, J Comput Chem 28 (2007) 2075-2084.
[129] Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E., Berendsen H.J., GROMACS: fast, flexible, and free, J Comput Chem 26 (2005) 1701-1718.
[130] Hubbard S.J., Thornton J.M., 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London., 1993.
[131] Pollastri G., Przybylski D., Rost B., Baldi P., Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles, Proteins 47 (2002) 228-235.
[132] Fuchs P.F., Alix A.J., High accuracy prediction of beta-turns and their types using propensities and multiple alignments, Proteins 59 (2005) 828-839.
[133] Bystroff C., Thorsson V., Baker D., HMMSTR: a hidden Markov model for local sequence-structure correlations in proteins, J Mol Biol 301 (2000) 173-190.
[134] Offmann B., Tyagi M., de Brevern A.G., Local Protein Structures, Current Bioinformatics 3 (2007) 165-202.
[135] de Brevern A.G., Camproux A.-C., Hazout S., Etchebest C., Tuffery P., Protein structural alphabets: beyond the secondary structure description, in: Sangadai S. (Ed.), Recent Research Developments in Protein Engineering, Research Signpost, Trivandrum 2001, pp. 319-331.
[136] Tyagi M., Benros C., Martin J., de Brevern A.G., Description of the local protein structure. II. Novel approaches., in: de Brevern A.G. (Ed.), Recent Research Developments in Protein Engineering, Research Signpost, Trivandrum 2007, pp. 34-47.
[137] Benros C., Martin J., Tyagi M., de Brevern A.G., Description of the local protein structure. I. Classical approaches, in: de Brevern A.G. (Ed.), Recent Research Developments in Protein Engineering, Research Signpost, Trivandrum 2007, pp. 1-33.
[138] de Brevern A.G., Etchebest C., Hazout S., Bayesian probabilistic approach for

predicting backbone structures in terms of protein blocks, Proteins 41 (2000) 271-287.
[139] de Brevern A.G., New assessment of a structural alphabet, In Silico Biol 5 (2005) 283-289.
[140] de Brevern A.G., Valadie H., Hazout S., Etchebest C., Extension of a local backbone description using a structural alphabet: a new approach to the sequence-structure relationship, Protein Sci 11 (2002) 2871-2886.
[141] de Brevern A.G., Benros C., Gautier R., Valadie H., Hazout S., Etchebest C., Local backbone structure prediction of proteins, In Silico Biol 4 (2004) 381-386.
[142] Etchebest C., Benros C., Hazout S., de Brevern A.G., A structural alphabet for local protein structures: improved prediction methods, Proteins 59 (2005) 810-827.
[143] de Brevern A.G., 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 32 (2007) 51-70.
[144] de Brevern A.G., Hazout S., Compacting local protein folds with a "hybrid protein model", Theo Chem Acc 106 (2001) 36-47.
[145] de Brevern A.G., Hazout S., 'Hybrid protein model' for optimally defining 3D protein structure fragments, Bioinformatics 19 (2003) 345-353.
[146] Benros C., de Brevern A.G., Etchebest C., Hazout S., Assessing a novel approach for predicting local 3D protein structures from sequence, Proteins 62 (2006) 865-880.
[147] Benros C., de Brevern A.G., Hazout S., Analyzing the sequence-structure relationship of a library of local structural prototypes, J Theor Biol 256 (2009) 215-226.
[148] Fourrier L., Benros C., de Brevern A.G., Use of a structural alphabet for analysis of short loops connecting repetitive structures, BMC Bioinformatics 5 (2004) 58.
[149] Etchebest C., Benros C., Bornot A., Camproux A.C., de Brevern A.G., A reduced amino acid alphabet for understanding and designing protein adaptation to mutation, Eur Biophys J 36 (2007) 1059-1069.
[150] Faure G., Bornot A., de Brevern A.G., Protein contacts, inter-residue interactions and side-chain modelling, Biochimie 90 (2008) 626-639.
[151] Tyagi M., Gowri V.S., Srinivasan N., de Brevern A.G., Offmann B., A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications, Proteins 65 (2006) 32-39.
[152] Tyagi M., Sharma P., Swamy C.S., Cadet F., Srinivasan N., de Brevern A.G., Offmann B., Protein Block Expert (PBE): a web-based protein structure analysis server using a structural alphabet, Nucleic Acids Res 34 (2006) W119-123.
[153] Tyagi M., de Brevern A.G., Srinivasan N., Offmann B., Protein structure mining using a structural alphabet, Proteins 71 (2008) 920-937.
[154] Dong Q., Wang X., Lin L., Prediction of protein local structures and folding fragments based on building-block library, Proteins 72 (2008) 353-366.
[155] Thomas A., Deshayes S., Decaffmeyer M., Van Eyck M.H., Charloteaux B., Brasseur R., Prediction of peptide structure: how far are we?, Proteins 65 (2006) 889-897.
[156] Dudev M., Lim C., Discovering structural motifs using a structural alphabet: application to magnesium-binding sites, BMC Bioinformatics 8 (2007) 106.
[157] Li Q., Zhou C., Liu H., Fragment-based local statistical potentials derived by combining an alphabet of protein local structures with secondary structures and solvent accessibilities, Proteins (2009).
[158] Zimmermann O., Hansmann U.H., LOCUSTRA: accurate prediction of local protein structure using a two-layer support vector machine approach, J Chem Inf Model 48 (2008) 1903-1908.
[159] Karchin R., Cline M., Mandel-Gutfreund Y., Karplus K., Hidden Markov models that use predicted local structure for fold recognition: alphabets of backbone geometry,

Proteins 51 (2003) 504-514.
[160] DeLano W.L.T., The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org (2002).
[161] Berendsen H.J.C., van der Spoel D., van Drunen R., GROMACS: A messagepassing parallel molecular dynamics implementation, Comp Phys Comm 91 (1995) 43-56.
[162] Humphrey W., Dalke A., Schulten K., VMD - Visual Molecular Dynamics, J. Molec. Graphics 14 (1996) 33-38.
[163] VMD, http://www.ks.uiuc.edu/Research/vmd/.
[164] Hollup S.M., Salensminde G., Reuter N., WEBnm@: a web application for normal mode analyses of proteins, BMC Bioinformatics 6 (2005) 52.
[165] Reuter N., Hinsen K., Lacapere J.J., Transconformations of the SERCA1 Ca-ATPase: a normal mode study, Biophys J 85 (2003) 2186-2197.
[166] Valadie H., Lacapcre J.J., Sanejouand Y.H., Etchebest C., Dynamical properties of the MscL of Escherichia coli: a normal mode analysis, J Mol Biol 332 (2003) 657-674.
[167] Hinsen K., The molecular modeling toolkit: A new approach to molecular simulations, J Comp Chem 21 (2000) 79-85.
[168] Hinsen K., Petrescu A., Dellerue S., Bellissent-Funel M., Kneller G., Harmonicity in slow protein dynamics, Chem.Phys 261 (2000) 25-37.
[169] 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 105 (2001) 6507-6514.
[170] Crump M.P., Gong J.H., Loetscher P., Rajarathnam K., Amara A., Arenzana-Seisdedos F., Virelizier J.L., Baggiolini M., Sykes B.D., Clark-Lewis I., 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 16 (1997) 6996-7007.
[171] Baysal C., Atilgan A.R., Elucidating the structural mechanisms for biological activity of the chemokine family, Proteins 43 (2001) 150-160.
[172] Skelton N.J., Quan C., Reilly D., Lowman H., Structure of a CXC chemokine-receptor fragment in complex with interleukin-8, Structure 7 (1999) 157-168.
[173] Comeau S.R., Gatchell D.W., Vajda S., Camacho C.J., ClusPro: a fully automated algorithm for protein-protein docking, Nucleic Acids Res 32 (2004) W96-99.
[174] Comeau S.R., Gatchell D.W., Vajda S., Camacho C.J., ClusPro: an automated docking and discrimination method for the prediction of protein complexes, Bioinformatics 20 (2004) 45-50.
[175] Fernandez-Recio J., Totrov M., Abagyan R., ICM-DISCO docking by global energy optimization with fully flexible side-chains, Proteins 52 (2003) 113-117.
[176] ICM, http://www.molsoft.com/.
[177] Hoogewerf A.J., Kuschert G.S., Proudfoot A.E., Borlat F., Clark-Lewis I., Power C.A., Wells T.N., Glycosaminoglycans mediate cell surface oligomerization of chemokines, Biochemistry 36 (1997) 13570-13578.
[178] Plater-Zyberk C., Hoogewerf A.J., Proudfoot A.E., Power C.A., Wells T.N., Effect of a CC chemokine receptor antagonist on collagen induced arthritis in DBA/1 mice, Immunol Lett 57 (1997) 117-120.
[179] Kuschert G.S., Coulin F., Power C.A., Proudfoot A.E., Hubbard R.E., Hoogewerf A.J., Wells T.N., Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses, Biochemistry 38 (1999) 12959-12968.
[180] Ali S., Palmer A.C., Fritchley S.J., Maley Y., Kirby J.A., Multimerization of monocyte chemoattractant protein-1 is not required for glycosaminoglycan-dependent transendothelial chemotaxis, Biochem J 358 (2001) 737-745.

[181] Singh S.K., Hora R., Belrhali H., Chitnis C.E., Sharma A., Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain, Nature 439 (2006) 741-744.

[182] Hans D., Pattnaik P., Bhattacharyya A., Shakri A.R., Yazdani S.S., Sharma M., Choe H., Farzan M., Chitnis C.E., Mapping binding residues in the Plasmodium vivax domain that binds Duffy antigen during red cell invasion, Mol Microbiol 55 (2005) 1423-1434.

[183] McHenry A.M., Adams J.H., The crystal structure of P. knowlesi DBPalpha DBL domain and its implications for immune evasion, Trends Biochem Sci 31 (2006) 487-491.