

3D-structural models of transmembrane proteins.

Alexandre G. de Brevern ^{1*}

¹INSERM UMR-S 665, DSIMB, Institut National de Transfusion Sanguine (INTS),
Université Paris Diderot - Paris 7, 6 rue Alexandre Cabanel, 75739 Paris Cedex 15, France.

* Corresponding author:

mailing address: Dr. de Brevern A.G., INSERM UMR-S 665, DSIMB, Institut National de
Transfusion Sanguine (INTS), Université Paris Diderot - Paris 7, 6 rue Alexandre Cabanel,
75739 Paris Cedex 15, France.

E-mail : alexandre.debrevern@univ-paris-diderot.fr

Tel: (33) 1 44 49 30 38

Fax: (33) 47 34 74 31

i. Summary / abstract

Transmembrane proteins are macromolecules implicated in major biological process and diseases. Due to their specific neighborhood, few transmembrane protein structures are nowadays available. The building of structural models of transmembrane proteins is a major research area. Due to the lack of available 3D structures, automatic homology modeling is not an efficient way to propose pertinent structural models. Hence, most of the structural models of transmembrane proteins are done through a more complex protocol. This latter comprises the use of secondary structure prediction to complete the comparative modeling process. Then, refinement and assessment steps are performed go often to a novel comparative modeling process. Nowadays, it is also possible to take attention to the helix – helix and helix – lipid interactions, and, to build even quaternary structures. In all cases, the taking into account of experimental data is the most important factor to proceed to correct structural models.

ii. Key words

Transmembrane protein structures, multiple sequence alignment, structural models, comparative modeling, homology modeling, secondary structure prediction, helix – helix interaction, helix – lipid interaction, bilayer membrane, protein docking.

1. Introduction

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 as receptors, transporters or channels. They are embedded in lipid membrane that constitutes a very specific neighboring. Due to this particularity, the obtaining of experimental 3D transmembrane structure is a hard case. The total number of transmembrane proteins in the Protein DataBank[1] is limited, comprising ~1% of available structures[2]. The design of structural models becomes an important axis of research. Indeed, more than 2/3 of the marketed drugs targets a transmembrane protein and 50% specifically a GPCR[3]. Thus, most of the time is not possible to work with an experimental structure, and, so 3D structural models are an important research fields for understanding biological mechanisms and interactions[4].

We present in this chapter the classical pipeline to build 3D structural models. The most common way to propose 3D structural models is based on a comparative modeling process coupled with transmembrane segment predictions. Nonetheless, the principle goes far beyond the classical homology modeling as often the target structure is not directly related to the query sequence, *i.e.*, it is not possible to simply align the sequence of the protein queries and targets. Thus, it is an iterative process mainly based on –when possible- multiple sequence alignments, bibliographic and web researches, molecular refinement, helix – helix and helix – lipid interaction prediction tools. Two papers can be read to have pertinent examples of such approach [5,6].

The chosen structural models must encompass most of the biochemical features and reflect the known experimental data. They may be used to analyze functional interaction properties.

2. Materials

A recent correct computer with internet is sufficient. Numerous software programs are available on-line, but some must be locally installed. Most of these software programs can be used with Windows Operating System (OS) while few works only with Linux OS.

Tables 22.1 to 22.3 summarize some of the available tools that can be used, *e.g.*, the secondary structure and interaction prediction methods. Updated and additional links could be found at <http://www.dsimb.inserm.fr/~debrevern/TM/index.html>.

3. Methods

The main principle of the approach is described in Figure 22.1. A first step consists in the true knowledge of the protein of interest (POI). Three complementary approaches are important: multiple sequence alignment, bibliographic and on-line research, and, secondary structure prediction. Then, the alignment of the POI (*query*) sequence with a (*target*) sequence of a protein with an available (*template*) 3D structure must be done. This information is sufficient to build protein structures thanks to comparative modeling approach. Different structural models are obtained. They must be refined and evaluated. Corrections must so be done to the original alignment until a structural model is found interesting. Then, many approaches can be performed to go deeper into the protein structural models (see the other chapters).

3.2 Sequence analyses: multiple sequence alignment and bibliographic researches.

1. The correct sequence of the (*query*) POI must be found. When alternative splicing or mutants are common, a careful attention must be taken to select the right sequence. The expasy server (Uniprot) is a famous web server as it is highly manually curated.

[Table 22.1 here]

2. Related protein sequences can be found thanks to PSI-BLAST. In a first step, this search is performed using only the Protein DataBank. If one hit appears with a correct sequence identity, it must be used as the structural template. Often, it is not the case.

3. In this second step, related sequences must be found to analyze evolutionary information. The search is so done on the Uniprot and on non-redundant databank of NCBI. As PSI-BLAST can take very redundant sequences and also select short sequences, the researcher must carefully select interesting proteins. The NCBI web service of PSI-BLAST is well conceived for such pruning schema (*see Note 1*).

4. As PSI-BLAST has been not created to optimally align sequences, but to search related sequences to a query one. The selected related proteins must then be aligned into a multiple alignment with dedicated software like Clustalw, T-coffee or Muscle. It allows locating conserved and non conserved regions of the protein sequences, i.e., more selective pressure can be expected to conserved regions. They are important for the protein fold / function. At the opposite, non conserved regions can be potentially more flexible. A phylogenetic approach -with a dendogram- is an efficient way to analyze the raw data (*see Note 2*).

5. A bibliographic research on the POI is obvious, but it is also important to search information on related protein sequences. In the same way, a good knowledge of related protein folds is essential. For this purpose, dedicated transmembrane protein websites, e.g., Stephen White, PDBTM and OMP, have a lot of pertinent information, both for the specialists and non-specialists. If the protein is a GPCR, GPCRdb is also an interesting tool to obtain information of this kind of protein.

6. Biological and experimental data are essential to build pertinent structural models. Summarize the biological data is one Table with the amino acid position, its kind and the corresponding biological property(ies) (*e.g.*, see Table 1 of [5]) (*see Note 3*).

3.3 Of helices and transmembrane segments.

[Table 22.2 here]

1. Table 22.2 summarizes different and recognized transmembrane secondary prediction methods. These web services are based on different approaches (hydrophobicity, sequence alignment, artificial neural networks ...) and on different protein structure databanks (ranging from the oldest one to the most recent ones). The principle is to put the sequence in a form and wait for the results.

2. Results can be given in very different ways ranging from the simple delimitation of the transmembrane repetitive secondary structure to the localization of extra- and intra-cellular loops and even a confidence index to the quality of the predictions (*see Note 4*).

3. As the number of available transmembrane protein structure is limited the predictions rate are to be taken carefully when no homologous sequences is associated to a known structure. Summarize the result is one Table (*e.g.*, see Table 2 of [5]). The high number of different approaches allows finding average helix number and positions; they will serve to start the comparative modeling. Bibliographic data could also eliminate a method which seems inadequate for this protein, *e.g.*, a method which misses half of the helices. In the same way, the confidence index can explain quantitatively why some regions are predicted very differently by the prediction methods. Indeed, some regions are *hard-to-predict* and require a special care (*see Note 5*).

4. An interesting point is to predict the secondary structure of the target structure, *i.e.*, the structure of the sequence that will be aligned with the query sequence. It permits to look at the behaviors of each secondary structure prediction methods.

3.4 Comparative modeling and other approaches.

1. Some web services able to propose structural models are available, *e.g.*, Homer,

consensus, Proteus 2 or HOMA. As the other web services, the sequence must be put in a form. Numerous methods are not dedicated to transmembrane proteins and cannot find correct transmembrane models. Nonetheless, they must be tried as they can give some hints about the compatibility folds.

2. Comparative modeling can be done with Modeller software. This latter is the most known and most used homology modeling software. It used an alignment between one *query* and one *target* sequence; the target sequence must have a known available (*template*) structure. The crucial part is so the quality of the alignment. Concerning our purpose, it is also the converging of repetitive secondary structure. Figure 22.2 summarizes the principle. (a) The POI (*query sequence*) must be cut in transmembrane domain, *i.e.*, the region with the transmembrane repetitive secondary structure (TRSS) and connecting loops, and the N and C termini regions. In paragraph 3.3, initial delineation of repetitive secondary structure has been performed. (b) The TRSSs of the target structure must be precisely located. (c) These TRSSs must be aligned with the corresponding TRSSs of the target structure. Same thing must be done for the loops, resulting in a global alignment of transmembrane domain. (d) The N and C termini regions can be considered as globular. So, classical homology modeling can be done with software of Table 22.2. Depending on the size and compatibility with related structures, it can be less or more complicated. (e) An entire alignment can so be done. Figure 22.3 shows an example of an alignment usable by Modeller (PIR format). This example is the alignment of (*target*) rhodopsin with (*query*) DARC.

3. Modeller needs this alignment, but also the corresponding template structure and a simple script which summarizes all these information. The template structure needs to be strictly equivalent to the *sequence* of the structure, *i.e.*, the sequence present in the PDB file. Indeed, it is classical to have not a complete equivalence of protein sequence and the resolved protein. For instance, the PDB structure of the rhodopsin (PDB code 1F88) has missing atoms

within a helix. The correspondence between sequence and structure of rhodopsin is not complete. Figure 22.4 gives an example of a Modeller script.

4. A series of models (at least 100) must be generated. It corresponds to the a.ending in the Modeller script. The recent Modeller software use Python language (*see Note 6*).

5. It is often interesting to explicitly add constraints on TRSSs, i.e., to add in the script the position of TRSSs to force the conservation of repetitive structure in the generated models. In the same way, constraints on known disulfide bridges or distances between residues lead to enhanced structural models. In Figure 22.4, the disulfide bond constraint has been added (class MyModel) for disulfide bonds 51-276 and 129-195 (*see Note 7*).

3.5 Structural models: assessment and refinement.

1. Modeller or equivalent approaches are quite good to define a correct topology. However, the structural models must be refined to obtain a better geometry. Thus, a first step often consists in selecting few models from the generated ones. It is possible to only look at the objective and DOPE functions given by Modeller for each of the generated models. Nonetheless, these functions can be used carefully as they were not fully tested with transmembrane proteins. Moreover, these functions can be highly sensitive to flexible regions. For instance, the DARC protein has a long flexible region [5], namely ECD1, these functions were only underlining the different conformations of this loop (only 20% of the protein). It was not discriminating (*see Note 8*).

2. Side-chains must be replaced thanks to a performing method, *e.g.*, SCWRL.

3. A minimization of the generated structural models is also a good requirement. The most powerful software today is Gromacs (see next chapter for details).

4. Then, A precise analysis of the models must be performed. Visualization of the protein structures with dedicated software, *e.g.*, rasmol, PyMol or VMD, is a first obligated

step.

5. Recognition of different types of errors in 3D models can be done by different software, *e.g.*, Verify 3D or Prosa. They are based on sequence – structure relationship statistics deduced from non-redundant databank. Moreover, no equivalent method is available for transmembrane proteins. Thus, only geometry of the structural models can be checked, *e.g.*, atom distances, angle values ... The most-used dedicated software are ProCheck and Whatcheck. It gives numerous values with summaries. It can highlight part of the structural models with wrong geometry or particular residues with a strange conformation. An important point is also to look at the target structure. Indeed, the template structure can have already some local un-canonical conformation.

6. The chosen structural models must encompass most of the biochemical features and reflect the known experimental data. If experimental data have shown that some residues are accessible, they must be found accessible in the selected models. Otherwise, the alignment must be corrected and the process of comparative modeling and analyses done again. For instance, the building of DARC [5] has needed 10 consecutive rounds with manual corrections of the alignment (*see Note 9*).

7. Other information can be used (see the paragraphs below).

3.6 Helix – helix interactions.

[Table 22.3 here]

1. Interaction between helices are well-described for globular proteins while for transmembrane some interesting works have been published. To have an idea of the interaction zones between helices is a major advantage to propose pertinent structural models. Table 22.3 shows the available methods of helix – helix and helix – lipid interactions.

2. A very elegant research [7] has shown the possibility to compute the compatibility

between two helices and deduced rules. Specific researches have quantified these compatibility statistics. In this limited field, kPROT is the most renowned approach. Nonetheless, it is not reachable for some times. Other approaches are not fully automated and so researchers need a strong background in bioinformatics to deal with this kind of approach.

3.7 Helix – lipid interactions.

1. These interactions are also essential for the protein folding and biological functions. It was less analyzed than helix - helix interactions, but nowadays, the number of available methods is higher than for helix – helix interactions. Often, *e.g.*, LIPS, they also provide hints to regions of interaction with other helices.

2. For LIPS, a multiple sequence alignment must be given, but the alignment concerns only one helix and must not contain one gap. So an important work is required to prepare the data.

3. The web server gives as a result not a single result, but the quantification of each possible face (7). Thus, a careful research of correspondence between these predictions and the selected structural models can be performed to improve the quality of the models.

3.8 Of loops and mutants.

1. Specific research can be done on extra- and intra-cellular loops. It is especially interesting in the case of GPCRs. Indeed, it is a common hypothesis that the 7 transmembrane helices have a strong conservation among its entire fold. This hypothesis is strongly supported by biochemical, biophysical and biological experiments. Hence, a major divergence between the GPCRs is the conformation of loops. A database like ArchDB can be use to select (manually) some alternative potential conformations. Then, a building of an alternative conformation can be generated thanks to comparative modeling.

The simplest approach is to use the selected models as the query and the template. In the target sequence, place gaps at the position of the loops. In the example of Figure 22.3, it could correspond to put two times the DARC_H1 sequence, one as the query sequence, one as the target sequence, and put the obtained structural models as the template. In the target sequence, put gaps at the position of the desired loop. Add a new sequence of the length of query sequence, but only with gaps. Add also the name of this sequence in the Modeller script (line sequence). At the position, change the gaps by the sequence of the loop. In the same way, add its PDB file into the script (to the line knows). It is possible to add constraints to all the rest of the protein. A molecular dynamics (or a simulated annealing) can help to analyze the flexibility of the loops (see next chapter). Recently, a novel class (MyLoop) in Modeller has been proposed to refine the loops, it is based on the optimization of Modeller's DOPE function.

2. Another potentiality of *in silico* building of transmembrane protein structural models is the proposition of “supervised” mutants. As the final selected models reflect the biological data, an analysis of the surface underlines the important positions. It is possible to mutate *in silico* the residues (a specific class exist in Modeller) and analyze a possible consequence of this mutation. A simple idea is to observe the electrostatic properties of the protein. Electrostatic potential mapped on the molecular surface can be coarsely done with Swiss-PdbViewer (aka DeepView) from SwissModel. It allows ranking of the mutation to be proposed.

3. Quaternary structures can be computed *in silico*. Two approaches exist: (i) from a known available complex, or (ii) by docking. For the first possibility, it consists -roughly- in the comparative / homology modeling of the different partners. The hypothesis is that the interaction regions re conserved and so must be also in the structural model complex. The second case is most common. If the protein complex is an homomer with a known symmetry,

it is a constraint that can be directly used during the comparative modeling approach. Otherwise, each partner must be build and docking approach must used. As for the building of one protein structural model, the building of a complete quaternary structure is greatly enhanced if biological data are included.

4. Notes

1. For the multiple sequence alignment, it is important to control the redundancy of the data as the length of the sequences. (i) a cluster of highly redundant sequences is not informative and bias the alignment analysis. (ii) PSI-BLAST often keep small sequence fragments, they are not pertinent in the case of a protein multiple sequence alignment.

2. With a pertinent multiple sequence alignment, it is also possible to locate important residues in related sequences. These positions can be used as biological constraints, the most important ones.

3. Summarize the data with Tables, for the prediction and for the important residues. Note than a mutation abolishing a binding, an interaction is not always directly at the binding site, but it can also be a crucial residue implicated in the maintaining of the fold (*Important is not always accessible*).

4. The prediction index is a great tool to analyze the difficulty of prediction. Some regions are clearly repetitive structures embedded in the bilayer or loop swimming in water while others are complicated. These latter must be carefully checked at each step of the process.

5. As the number of available protein structure is quite low, it is very important to analyze properly the known folds and the rules [8] that govern this kind of protein folds. However, do not follow blindness the common features of transmembrane proteins. For

instance, Tryptophan is known to be preferred at the lipid water interface, but often also it is not the case.

6. Multiple structural models must be done to find at least a pertinent one. In the same way, the testing of alternative alignment is an essential task. Another interesting approach is to predict also structural models of related proteins. In the same way, if one structural model of a related protein is available, it can be used as a structural template.

7. Proline and kinks in repetitive structures are always a problem. For instance, rhodopsin as 3 kinked helices, and the K⁺ channel one essential ones.

8. N and C termini regions of the transmembrane protein can be long fragments. In this case, it is possible to use classical methods available on the web, e.g., Proteus 2, and use the results as novel template only for one part of the template (see paragraph 3.8). In the same way, the use of multiple templates is an excellent tool to sample the potential conformations of the structural models.

9. As presented in this chapter, it must be noticed that the building of a pertinent transmembrane protein structural model is mainly a manual approach and that the most important constraint is the biological (experimental) one.

Acknowledgements. This work was supported by French Institute for Health and Medical Care (INSERM) and University Paris 7 – Denis Diderot.

5. References

- [1] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-42.
- [2] Fleishman, S.J., Unger, V.M. and Ben-Tal, N. (2006). Transmembrane protein structures without X-rays. *Trends Biochem Sci* 31, 106-13.
- [3] Klabunde, T. and Hessler, G. (2002). Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* 3, 928-44.
- [4] Radestock, S., Weil, T. and Renner, S. (2008). Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring. *J Chem Inf Model* 48, 1104-17.
- [5] de Brevern, A.G., Wong, H., Tournamille, C., Colin, Y., Le Van Kim, C. and Etchebest, C. (2005). A structural model of a seven-transmembrane helix receptor: the Duffy antigen/receptor for chemokine (DARC). *Biochim Biophys Acta* 1724, 288-306.
- [6] Hazai, E. and Bikadi, Z. (2008). Homology modeling of breast cancer resistance protein (ABCG2). *J Struct Biol* 162, 63-74.
- [7] Fleishman, S.J. and Ben-Tal, N. (2002). A novel scoring function for predicting the conformations of tightly packed pairs of transmembrane alpha-helices. *J Mol Biol* 321, 363-78.
- [8] von Heijne, G. (2007). The membrane protein universe: what's out there and why bother? *J Intern Med* 261, 543-57.

Figure Captions

Figure 22.1. Pipeline to build a structural model of a transmembrane protein structures. The protein sequence of interest is analyzed thanks to multiple sequence analysis, bibliographic and on-line researches, and, secondary structure prediction methods. Then a comparative modeling with constraints is performed. The structural models are refined and analyzed. It is an iterative process; the local modification of alignment gives new structural models which must be analyzed ...

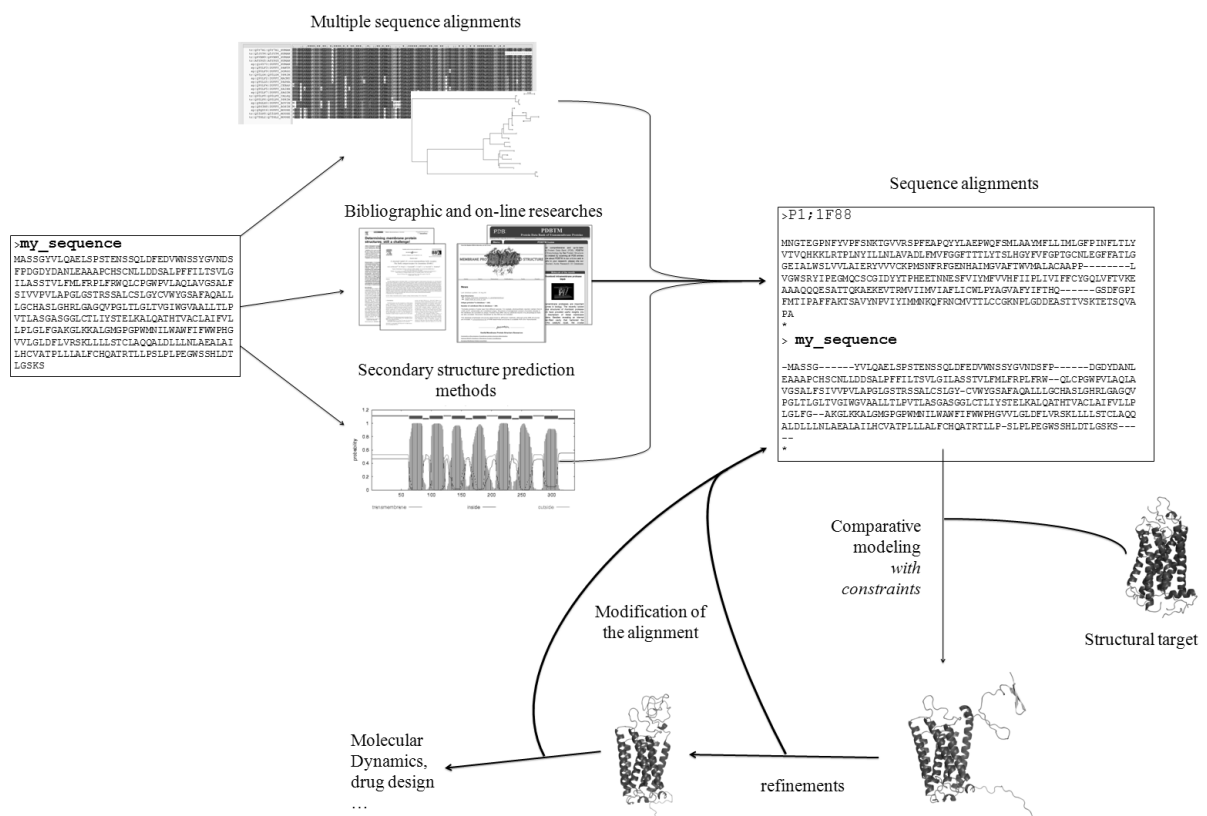


Figure 22.2. *Construction of the alignment of the transmembrane protein.* The protein sequence must be cut into 3 parts: transmembrane domain and the N and C termini regions. (a) The predicted repetitive structures are placed on the sequence. (b) The repetitive structures are assigned on the template structure. (c) The correspondence between predicted and assigned regions must be done. Each local alignment is optimized with dedicated software as Clustalw. (d) The N and C termini regions can be treated as globular proteins, specific predictions can so be done (or not). (e) A complete alignment is done corresponding to the entire protein.

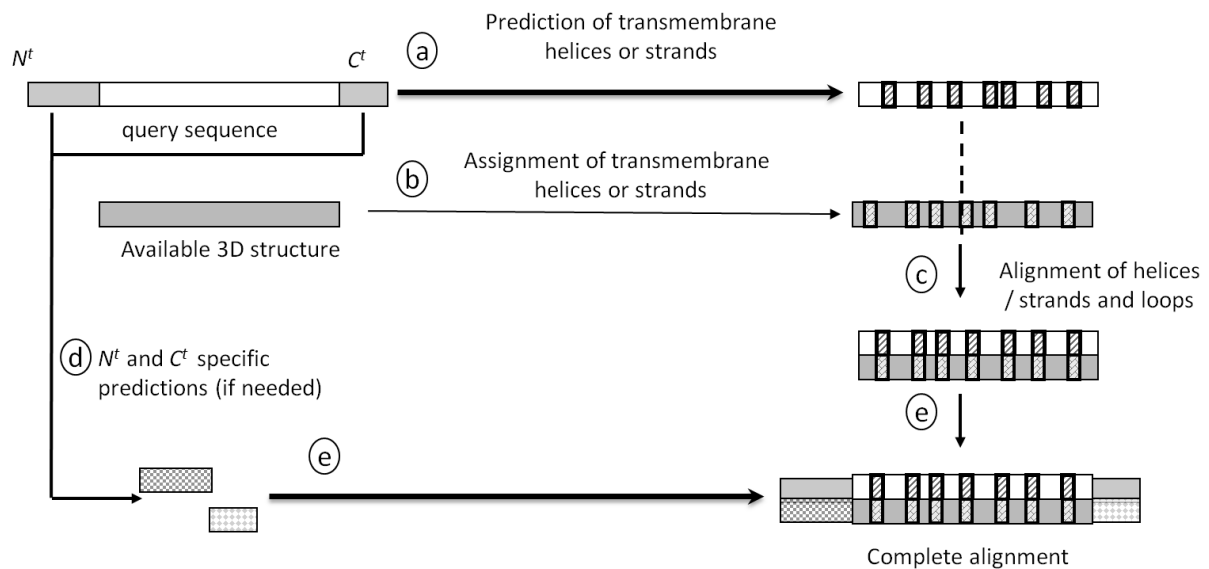


Figure 22.3. *Example of protein sequence alignment usable by Modeller.* The alignment was done between the sequence query DARC and the sequence of the target (structural template), the rhodopsin. The alignment file is named darc.ali.

```
>P1;1F88
structureX:1F88:  1 :A : 348 : :RHODOPSIN : BOS TAURUS:  2.80 :-1.0
----MNGTE-GP-----NFYVPFSNKTGVVRSPPFEAPQYYL---A-EP
WQ----FS--MLAAYMFLLIMLGFPINFLTLY-VTVQHKKLR-TP-LNYIL
LNLAV-ADLF-MVFGGFTTTLTYTSLHGYFVFGPTGC-NLE--GFFATLGGE
IALWLSLVLAIERVYVVCKPMSNFRFGENH---AIMGVAF-TWVMALACAA
PPLVGWSRYIPEGMQCSCGIDYYTPHEETNNE---SFVIYMFVVHFI IPLI
VIFFCYQLVFTVKEAAA----SATTQKAEKEVTRMVIIMVIAFLICWLPY
AGVAFYI--FTHQ-----GSDFGP-----IFMTI PAFFAKTSAVYNPVIYI
-MMNKQFRNCMVTTLCCKGNP-----
*
```

```
>P1;DARC_H1
sequence:DARC_H1:  1 : : 336 : : DARC HUMAN : : :
MGNCLHRAELSPSTENSSQLDFEDVWNSSYG-VNDSFPDGDYDANLEAAAP
CHSCNLLDDSDALP-FFILTSVLGILASSTVLFMLFRPLFRWQLCPGWP-VL
AQLAVGSALFSIVVPVLAPG-----LGSTRSSALCSLGYCVWYGSA
FA--QALLLG-----CHASLGHRLGAGQVPGLTLGLTVGIWGVAALLTL
P-VT-LASGASGGL---CTLIYSTEKLA----LQATHTVACLAIFVLLPLG
--LFG-AK---GLKKALGMGPGPW-----MNILWAWFIFWWPH
-GVVLGLDFLVRSKLLLLSTCLAQQALDL-LLNLAEALAILHCVATPLLLA
LFCHQATR-TLLPSL-----PLPEG-WSSHLDTLGSKS-----
*
```


Figure 22.4. *Example of Modeller script.*

```
# Homology modeling by the automodel class
from modeller import *          # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class

# Redefine the special_patches routine to include the additional disulfides
# (this routine is empty by default):
class MyModel(automodel):
    def special_patches(self, aln):
        # A disulfide between residues 51 and 276:
        self.patch(residue_type='DISU', residues=(self.residues['51'],
                                                    self.residues['276']))

        # A disulfide between residues 129 and 195:
        self.patch(residue_type='DISU', residues=(self.residues['129'],
                                                    self.residues['195']))

log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

a = automodel(env,
              alnfile = 'darc.ali',      # alignment filename
              knowns   = '1F88',         # codes of the templates
              sequence = 'DARC_H1')      # code of the target
a.starting_model= 1 # index of the first model
a.ending_model  = 100 # index of the last model
                  # (determines how many models to
                  # calculate)
a.make()          # do the actual homology modeling
```

Table Captions

Table 22.1. *Web services and softwares.*

Table 22.2. *Transmembrane repetitive structure prediction web services.*

Table 22.3. *helix – helix, helix – lipid and miscellaneous web services.*

Tables

Table 22.1. *Web services and softwares.*

name	purpose	soft.	url
Protein Databank	database of available protein structures	on-line	http://www.rcsb.org/pdb/home/home.do
Stephen White laboratory	transmembrane protein structures	on-line	http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html
UniProt KB / Swiss-Prot	protein sequences	on-line	http://www.expasy.ch/sprot/
PDBTM	information on transmembrane proteins	on-line	http://pdbtm.enzim.hu/
OMP	orientations of Proteins in Membranes (OPM) database	on-line	http://opm.phar.umich.edu/
GPCRdb	information on GPCR	on-line	http://www.gpcr.org/7tm/
PSI-BLAST	search protein database	Linux / on-line	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Modeller	homology modeling	Windows / Linux	http://salilab.org/modeller/modeller.html
Swiss model	homology modeling (and other)	Windows / Linux	http://swissmodel.expasy.org/SWISS-MODEL.html
Homer	homology modeling	on-line	http://protein.cribi.unipd.it/Homer/
Consensus	homology modeling	on-line	http://structure.bu.edu/cgi-bin/consensus/consensus.cgi
PROTEUS2	homology modeling	on-line	http://wishart.biology.ualberta.ca/proteus2
HOMA	homology modeling (and other)	on-line	http://www-nmr.cabm.rutgers.edu/HOMA/
clustalw	sequence alignment	all	http://www.ebi.ac.uk/Tools/clustalw2/index.html
T-Coffee	sequence alignment	Linux / on-line	http://www.ebi.ac.uk/Tools/t-coffee/
Muscle	sequence alignment	Linux / on-line	http://www.drive5.com/muscle/
rasmol	protein visualisation	Windows / Linux	http://www.rasmol.org/
vmd	protein visualisation	Windows / Linux	http://www.ks.uiuc.edu/Research/vmd/
pymol	protein visualisation	Windows / Linux	http://pymol.sourceforge.net/
WhatCheck	structural model assessment	Linux	http://swift.cmbi.ru.nl/gv/whatcheck/
ProCheck	structural model assessment	Linux	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
SCWRL	side-chain replacement	Linux / on-line	http://dunbrack.fccc.edu/SCWRL3.php
Gromacs	molecular modeling	Linux	http://www1.jcsg.org/prod/scripts/scwrl/serve.cgi http://www.gromacs.org/

Table 22.2. *Transmembrane repetitive structure prediction web services.*

web services	year	url
all-α		
TMbase	1993	http://www.ch.embnet.org/software/TMPRED_form.html
TopPred II	1994	http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=toppred
TMAP	1994	http://bioinfo4.limbo.ifm.liu.se/tmap/index.html
MEMSAT	1994	http://saier-144-37.ucsd.edu/memsat.html
PredictProtein (include PhDTm)	1995	http://www.predictprotein.org/
TSEG	1998	http://www.genome.ad.jp/SIT/tsegdir/
TM-Finder	1999	http://www.ccb.sickkids.ca/tools/tmfinder/cgi-bin/TMfinderForm.cgi
PRED-TMR 2	1999	http://athina.biol.uoa.gr/PRED-TMR2/
HMMTOP 2.0	2001	http://www.enzim.hu/hmmtop/
OrienTM	2001	http://o2.biol.uoa.gr/orienTM/
DAS-TMfilter (DAS 2.0)	2002	http://mendel.imp.ac.at/sat/DAS/DAS.html
SOSUI	2002	http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html
SPLIT 4.0	2002	http://split.pmfst.hr/split/4/
THUMBUP - UMDHMM ^{TMHP}	2003	http://sparks.informatics.iupui.edu/Softwares-Services_files/umdhmm.htm
Phobius (new TMHMM)	2004	http://phobius.sbc.su.se/
ConPred II	2004	http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/
PRODIV-TMHMM	2004	http://www.pdc.kth.se/~hakanv/prodiv-tmhmm/
SVMtm	2004	http://ccb.imb.uq.edu.au/svmtm/svmtm_predictor.shtml
TUPS	2005	http://sparks.informatics.iupui.edu/Softwares-Services_files/tups.htm
Localizome	2006	http://localodom.kobic.re.kr/LocaloDom/index.htm
MINNOU	2006	http://minnou.cchmc.org/
MEMSAT3	2007	http://bioinf.cs.ucl.ac.uk/memsat/
MemBrain	2008	http://chou.med.harvard.edu/bioinf/MemBrain/
all-β		
BDM	2002	http://gpcr.biocomp.unibo.it/biodec/
TBBpred	2004	http://www.imtech.res.in/raghava/tbbpred/
PROFtm	2004	http://www.rostlab.org/services/ProfTMB/index.html
BOMP	2004	http://www.bioinfo.no/tools/bomp
PRED-TMBB	2004	http://bioinformatics2.biol.uoa.gr/PRED-TMBB/index.jsp
TMBETA-DISC	2005	http://psfs.cbrc.jp/tmbetadisc/
TMBETA-SVM	2005	http://tmbeta-svm.cbrc.jp/
TMBETA-NET	2005	http://psfs.cbrc.jp/tmbeta-net/
TMB-HMM	2006	http://bmbpcu36.leeds.ac.uk/~andy/betaBarrel/TMB_Hunt_2/TMB_HMM.cgi
transFold	2006	http://bioinformatics.bc.edu/clotelab/transFold/

Table 22.3. *helix – helix, helix – lipid and miscellaneous web services.*

web services	year	url
<u>helix-helix</u>		
Helix Analysis Computational Tool	2002	http://bioinfo.tau.ac.il/~sarel/HelAna.html [not web service]
kprot	2004	not reachable
TMLIP-H	2005	not simply usable
<u>helix-lipid</u>		
ProperTM	2004	http://icb.med.cornell.edu/services/propertm/start
TMLIP-C	2005	not simply usable
LIPS	2006	http://gila.bioengr.uic.edu/lab/larisa/lips.html
TMX	2007	http://service.bioinformatik.uni-saarland.de/tmx/about.htm
<u>Misc.</u>		
TMRPres2D	2004	http://biophysics.biol.uoa.gr/TMRPres2D/
Archdb	2004	http://sbi.imim.es/cgi-bin/archdb/loops.pl
Triton	2008	http://ncbr.chemi.muni.cz/triton/