# Molecular Replacement techniques for high-throughput structure determination

Marc Delarue
Unite de Biochimie Structurale
Institut Pasteur
URA 2185 du C.N.R.S.
25 rue du Dr Roux
75015 Paris, France

## Introduction

In the context of structural genomics projects there are two main roads to consider for solving efficiently and rapidly the 3D structure of the target gene products by crystallography.

The first one is the MAD technique (reviewed in the next chapter of this volume...), which necessitates growing SeMet-substituted protein crystals, and the second one is Molecular Replacement, which demands X-Ray data for the native protein as well as the structure of a related homolog.

Molecular Replacement (MR) is an ensemble of techniques that aims at placing and orienting an approximate molecular model in the unit cell of the crystal being studied. This will provide the starting phases needed to calculate the initial electron density map in which the protein model can be built, either manually by iterative use of reconstruction with molecular graphics packages (1) followed by refinement (2), or even automatically if diffraction data up to 2.3 Angstroms or better are available (ARP/wARP (3), Solve/Resolve (4)...).

In this article, we will not focus on recent developments in refinement techniques, which benefited recently from better statistical treatments such as maximum likelihood targets for refinement (5), but rather will describe in detail some of the newest developments in MR to get the best possible set of phases to initialize refinement and reconstruction in the best possible conditions.

As more and more structures are deposited in the PDB (6), the chances of finding a (remote) homolog structure in the PDB have become higher and Molecular Replacement techniques are therefore more and more useful. This is reflected for instance in the number of hits one gets by doing a search for the keywords 'Molecular Replacement" in one of the leading journals in the protein crystallography community, e.g. Acta Cryst D (see Fig. 1).

Also, the growing number of sequences in the sequence databases have made the detection of remote homologs a more reliable task, as multi alignments of several dozens of sequences allow for building position-dependent mutation matrices (profiles) which are much more sensitive than pairwise comparisons with a standard mutation matrix (7). Having detected a homolog of known 3D structure, a model can be quickly derived using standard and automatic comparative modeling techniques (8). This may in turn result in an increase of the signal/noise ratio of Molecular Replacement searches.

One of the novelty in the field, which owes much to bioinformatics, is the apparition of a suite of programs that make use of all these new developments in a single web-interfaced suite of programs performing a multi alignment of a set of sequences and making use of all the homologs found in the PDB to scan many different models through the Molecular Replacement procedure (9). Refinement of the best solutions can also be performed to assess the correctness of the solution(s).

In addition, detection of remote structural homologs by threading methods can be made through, for instance, a suite of programs installed as a meta-server, which sends requests

to several web servers throughout the web and uses the results of each tasks to feed in the next request (10).
So there is a trend towards integrating different servers and/or packages in automatic protocols of MR, with the aim of obtaining the best possible model, which altogether greatly enhances the chances to find the solution quickly.

At the same time, it is apparent that the MR method itself continues to be developed and improved by a number of new ideas. This is in turn reflected in the number of hits of the same keywords "Molecular Replacement" in Acta Cryst A, a journal used mainly to describe methodological advances in crystallography: the cumulated number of hits between 1985 and 2004 is 534, an average of 27 articles per year. With the development of web-based interfaces, many of these new ideas are in fact implemented in programs that are available on-line (as well as their manual) and it has become more and more easy to test them rapidly, as they readily accept standard formats for the coordinates of the model and the X-Ray diffraction data.
We will review here most of these new methods as well as the "classical" ones and will mention their web-site address (see Table 1) with a brief mention of their main characteristics. We have tested all of them for the purpose of this review, using the same input data files, and will mention the used protocols throughout the text.

The rule of thumb for a successful application of molecular replacement is that the model should have an rmsd~2.0-2.5 Angstroms with the target structure, corresponding to a sequence identity with the target of 25-35%. In practice, however, there are many more structures solved by MR in the PDB using models with sequence identity of 60% or higher than otherwise.

There are really two issues in the application of molecular replacement automatic protocols to structural genomics:

i) Can one push the limits of the method so as to use models with less and less sequence identity with the target, i.e. models obtained by threading methods, with sequence identity levels between 15-25%? The answer seems to be yes (11), with newer methods being able to solve problems with sequence identity around 20% (9; see also G. Labesse, personal communication; Abstract in the GTBio Meeting, June 2004, Lyon, France).

ii) How far should one pursue efforts to solve the molecular replacement problem once the automatic protocols have failed; in other words, how much time should one spend in trying to solve difficult cases, before deciding to go back to the bench and growing SeMet crystals, or use SAD method with crystals soaked with one anomalous scatterer? This is actually a difficult question which depends on a lot of different issues, like the expertise already present in the lab in MR techniques, the solubility of the protein(s), which might or might not change upon selenomethionylation... This question cannot be answered in general but this is clearly an issue one should keep in mind in defining the strategy and automatic protocols of structure solution by MR.

## Test data used throughout this study with the different MR packages

The same data set at 2.5 Angstrom resolution, collected at the ESRF (ID14-EH2), for a pathogenic 6-phosphogluconolactonase (6PGL) (Delarue, Haouz and Stoven, unpublished data) was used throughout this study. The cell parameters are 70.3, 80.8, 90.3 in P2(1)2(1)2(1). There are two molecules in the asymmetric unit, related by a pseudo-translation, discovered by sorting the native Patterson peaks (see Protocol #1).
There are two possible models: 6PGL from *T. maritima* (1PBT) (about 40% sequence identity) and the glucosamine-6-phosphate deaminase (1DEA), which was detected by BLAST (25% sequence identity).

Protocol # 1: Checking your data

- check for completeness and redundancy at the desired resolution: usually a complete data set at 10-4 Angstroms is fine.
- check for possible twinning; read carefully the output of TRUNCATE in CCP4 (12) or submit your data to the scrutiny of Todd Yeates web site (http://www.doe-mbi.ucla.edu/Services/Twinning)
- check your space group (!)
  - for instance, if P2(1)2(1)2(1), check extinctions carefully
  - if polar space group, remember to try both enantiomorphic possibilities in the translation search, because they cannot be distinguished through Patterson methods used in the rotation function: P4(1) and P4(3), P6(1) and P6(5)...
- Check for possible Non Crystallographic Symmetry (NCS):
  - check for a possible pure translational non-crystallographic symmetry (pseudo-symmetry) by calculating a native Patterson function and sorting its peaks.
  - check for 2-fold axes, 3-fold axes... using self-rotation function.
  - calculate the percentage of solvent in the crystal, for different hypothesis as to the number nmol of molecules in the asymmetric unit: 100*(1. - nmol*(MW/Vm)/ Va.u.) where MW is the Molecular weight of a monomer, Vm is the density of a protein (~0.73 g/cm3) and Va.u. is the volume of the asymmetric unit. The percentage of solvent should be in the range 20-80%.

Remark: It might be of interest to read in detail a recent “tour de force” success story using Molecular Replacement in a very difficult case, with both high NCS (12 copies in the asymmetric unit) and twinning of the data (as discovered quite late in the process of structure solution): see S. Lee, M. Sawaya and D. Eisenberg (13).

---

## The standard Molecular Replacement method

## Historical background: Patterson methods.

The possibility and feasibility of Molecular Replacement was demonstrated by Rossmann and coll. in the 60's, as part of an effort to use non-crystallographic symmetry to solve the phase problem for macromolecules (14).

In this article, we will restrict ourselves to the case of finding the orientation and position of a known model in another unit cell, to help solve the molecular structure contained in this unit cell (see Fig. 2).

Traditionally, this 6-dimensional search (3 orientation angles and 3 translations are to be found) is divided into 2 separate and consecutive 3D search problems.

The first one consists in finding the orientation of the model through the so-called rotation function, which is defined in such a way that it should be maximum for the true orientation. In its real-space formulation, it consists in the convolution of the experimental Patterson function with the computed Patterson of the model in every possible orientation. If restricted to the molecular volume, the use of Patterson functions allows the superposition of intra-molecular vectors, which are independent of relative translations between the model and the target. In its reciprocal-space formulation (15), it was demonstrated that the maximum of this function is indeed the expected solution. Later on, the formula was rearranged using the plane-wave expansion and spherical harmonics so as to use the powerful technique of FFT (16), whose numerical implementation was further refined and stabilized by Navaza (17).

The second step consists in calculating a convolution of interatomic vectors between symmetry-related molecules of the correctly oriented model placed at different origins, with the experimental Patterson function. The reciprocal version of the calculation lends itself particularly well to Fourier (18) and FFT techniques and is in general very rapid.

Usually, the top peaks of the translation search are then submitted to a low resolution quick rigid-body refinement, for which quick algorithms have been devised (19,20).

The resolution is usually taken to be 12-4 Angstroms or so; if one wants to use the low resolution terms, one should use a solvent effect correction technique (21).

The following issues are crucial for the success of the method:

-completeness and quality of the data
-accuracy of the model(s)
-completeness of the modeling of the unit cell
-type of score used as an indicator of the quality of the agreement.

This is reflected in the following input parameters:

-quality of the X-Ray data (see Protocol #1)
-quality of the model(s) : e.g. poly-Ala vs full model (see below, structural diversity)
-number of molecules per asymmetric unit
-target functions

While the first point was addressed directly in Protocol #1, we will address the other points successively in the next paragraphs.

Before executing the real MR protocol, we strongly recommend to run a test case. Understanding the conventions of the rotation function will be greatly facilitated by reading a recent article (22) and the following protocol.

Protocol #2: Test case on the same model, the same space group, with calculated data

1. Rotate model by kappa around axis given by (phi,psi) using the following jiffy code:

```
       ck=cosd(kappa)
      sk=sind(kappa)
      cp=cosd(psi)
      sp=sind(psi)
      cf=cosd(phi)
      sf=sind(phi)
      a(1,2)= cp*sk+cf*sf*sp*sp*(1-ck)
      a(2,1)=-cp*sk+cf*sf*sp*sp*(1-ck)
      a(1,3)=-sf*sp*sk+cf*cp*sp*(1-ck)
      a(3,1)= sf*sp*sk+cf*cp*sp*(1-ck)
      a(2,3)= cf*sp*sk+sf*cp*sp*(1-ck)
      a(3,2)=-cf*sp*sk+sf*cp*sp*(1-ck)
      a(1,1)=          ck+cf*cf*sp*sp*(1-ck)
      a(2,2)=          ck+sf*sf*sp*sp*(1-ck)
      a(3,3)=          ck+          cp*cp*(1-ck)
c...   now rotate and translate, knowing the coordinates of the center of gravity xg
      do 100 k=1,3
            xnew(k) = xg(k) + a(k,1)*(x(1)-xg(1)) + a(k,2)*(x(2)-xg(2)) + a(k,3)*(x(3)-xg(3))
100    continue
```

2. Calculate Structure Factors (by CCP4 sfall) for the transformed coordinates **xnew** *in your own space group.*

3. Run your favorite molecular replacement program using i) the unrotated model as a search model and ii) calculated data as experimental data.

4. Analyse results: make sure you understand the symmetry of the rotation function group, the translation solution (the y coordinate is undetermined in P2(1) etc...).
Get a feeling of the (maximum) height of the signal you can expect.

5. Convince yourself that (kappa, phi, psi) applied in 1. is the same as the solution of MR.

```
c...   Hint: here is the rotation matrix using eulerian angles alpha, beta, gamma (see Ref. 22)
c...   Caveat: AMoRe first rotates the model so that the inertial axes coincide with x, y and z.
      ca=cosd(alpha)
      sa=sind(alpha)
      cb=cosd(beta)
      sb=sind(beta)
```

```
cg=cosd(gamma)
sg=sind(gamma)
mat(1,1)=-sa*sg + ca*cb*cg
mat(2,2)= ca*cg - sa*cb*sg
mat(3,3)= cb
mat(1,2)= ca*sg + sa*cb*cg
mat(2,1)=-sa*cg - ca*cb*sg
mat(1,3)=-sb*cg
mat(3,1)= ca*sb
mat(2,3)= sb*sg
mat(3,2)= sa*sb
```

---

Here we show a typical command file for the AMoRe package (23), one of the most widely used packages of MR.

Protocol # 3: A typical Molecular Replacement session using Amore (23)

1. Read the Manual (!)

2. Move to an empty directory and type : *csh $AMORE/setup*

3. Then

- in ./d/:
  - create data.d (as data.example created by setup), and give cell parameters, symmetry related positions, resolution limits and the number of molecule(s) (and their type) to search in the asymmetric unit

```
 ** 6PGL ** Title
70.315   80.852   90.31   90. 90. 90.
x,y,z * 1/2+x,1/2-y,-z * 1/2-x,-y,1/2+z * -x,1/2+y,1/2-z * end
0               ; orthogonalising code
95.   0.0       ; % reflections, B-add
12.   4.0       ; resolution range
1   2           ; NTYP, (nmol(n),n=1,NTYP)
```

  - data files with diffraction data and model coordinates must be named hkl.d and xyzN.d ... , respectively; N=1 for the first model, N=2 for the second one...
  - insert FORMAT card (upper case) in hkl.d and xyz{#}.d files (as hkl.example and xyz1.example created by setup).
- in ./i/:

  edit dato.i3 file which will look like this:

```
job  +*+*+*+*+*+*+*+*+*
xyz                  model type (could ne a map)
1.   2  2  0.5  2.             rot: %rad, lmin's, rotation function cutoff, step in degrees
c-c   50 0.5  50   1-body trans: option, nb. orient. to try, cutoff, nb. peaks
p-t   10 0.5  50   n-body trans: option, nb. orient. to try, cutoff, nb. peaks
```

```
10    20                    fitting: nb. trans. to fit, nb. of iterations
0.                        packing: CoM cutoff  for -crude- packing function
```

4. Finally, in ./:  
   *csh ./e/job dato*  
   *csh ./job* (protocol created by ./e/job for automated runs).

5. In our test case (6PGL), there was one solution well-detached with a correlation coefficient of about 30%. However, we were unable to bring the R-factor below 52%, even after refinement, until we collected another data set with crystal soaked in a mercury derivative. This derivative turned out to be highly isomorphous and refinement with CNS produced an R-factor around 48%. Phases from the model were able to pull out the heavy atom sites through both isomorphous and anomalous Fourier difference maps. The resulting solvent-flattened 2.8 Angstrom SIRAS map was interpretable.

---

### How does one know the solution is found?

Sometimes, it is not so easy to convince oneself that the solution of the molecular replacement solution has in fact been found, even after rigid-body refinement; indeed, the first solution is not always well detached and different scores produce different rankings. The most commonly used scores are correlation coefficients on either intensities or structure-factor amplitudes, and R-factors. Even though these criteria are formally related (24), they can produce different rankings especially if no solution is clearly detached. Some other criterion is then needed to discriminate between the potential solutions.

One possibility is to run simulated annealing refinement in torsion angle space as implemented in CNS (25). As this is one of the most powerful program in terms of radius of convergence, it is especially useful to look for the decrease of the free-Rfactor (5), but this is a rather cpu-intensive task if several possible solutions are to be tested.

In difficult cases, there is actually a way to bring the R-work and R-free down by other means than just rigid-body or dihedral angle molecular dynamics, while still exploring just a few degrees of freedom, using the Normal Mode Analysis protocol (26). This is implemented in http://lorentz.immstr.pasteur.fr and will be described in the paragraph on “exploring structural diversity” (see below).

Another possibility is to look at the packing of the top 10 solutions, because this can prove a very discriminating criterion. Even if the solution is well detached, it is still mandatory (and reassuring) to examine carefully the corresponding packing arrangement. This necessary (but not sufficient) solution can actually be implemented in the translation function using analytical expression that can be evaluated using FFT (27,28) and this is available in several MR packages, such as AMoRe (23).

## The need for finding a better score for the rotation function; the Phaser euphoria

Many people recognized that the rotation function suffers from some drawbacks and tried to improve the score by using origin-removed Patterson functions, normalized structure factors E-values... (29).

Brunger and coll. developed a “direct rotation function”, which is just a correlation coefficient between $E_{obs}$ and $E_{mod}$(Omega), the normalized structure factors of the crystal and of the rotated model, respectively. However, in this case the model is placed in a P1 cell of dimensions and angles identical to the ones of the crystal being studied. This works well but requires quite a lot of cpu as it is not amenable to FFT (30).

Another idea is to use different runs of the same program with slightly different models; it is most aptly described as an application of a “consistency principle”, namely it is required that the solution should appear consistently in all runs, even with a rather low score. Special algorithms have been developed to cluster similar solutions in eulerian angles space and convincing results have shown that it is indeed possible to increase the signal-to-noise ratio of the rotation function in this way (31).

Recently, R. Read extended the maximum likelihood formalism to the molecular replacement problem. Maximum likelihood puts on analytical grounds the notion that the best model is the one which maximizes the probability of having measured the actual experimental data at hand (structure factors). In his first implementation (BEAST), R. Read could show that he would get a much clearer signal in difficult cases, but the program was pretty slow (32). Recently, he and his team developed a much more rapid version of his algorithm (PHASER) based on an approximation which can be evaluated by FFT and this was incorporated in a single package that also contains the translation function (33). Roughly speaking, the rotation score is now based on a *weighted* origin-removed convolution of experimental and calculated Patterson functions. However, the authors stress that their program does not work very well in case where pseudo-translational symmetry is present. This is because the formalism assumes that the translation vector between the molecules in the unit cell is randomly distributed so that the relative phases between pairs of symmetry-related molecules are sampled randomly. This assumption is obviously violated if there is pseudo-translational symmetry; however, the authors mention that they have found a way around this problem and that it will be fixed soon.

Once released, the program got immediate praise from the crystallographic community, with thanking messages posted to the CCP4 Bulletin Board, reporting how several structures that had resisted molecular replacement traditional methods for years got solved in less than a day by PHASER.

So this is certainly one the packages to try first when dealing with a tough MR problem.

## Screening many solutions (multi sunt vocati, pauci vero electi)

The need for automated protocols is already apparent from the strategy adopted by AMoRe to circumvent the problem that the score of the rotation function (RF) is far from being perfect and does not always rank the solutions correctly (23). Indeed, it is often observed that the true solution is not the top solution, with many false positives. Hence, AMoRe runs a translation function (TF) for each of typically the top 50 or 100 solutions of the rotation function. This is actually quite rapid as TF is based on FFT; then, the first 10 solutions of each of these TF runs is in turn refined using a very effective implementation of rigid-body refinement (23).

One complication occurs with polar space groups, where all the possibilities must be tried in the TF. If there is an ambiguity in the extinctions, for instance in the Laue group Pmmm, again all possibilities must be searched. This is usually done ‘by hand”, going through all different possibilities one by one.

## NCS protocols and 6-D search programs

If there is NCS in the crystal, all molecules of the asymmetric unit must be searched in turn; every time a potential solution has been found, it is possible to use this information to increase the signal-to-noise ratio of the searches for the other molecules. But then, the combinatorics of testing the 50 top solutions of the rotation function and then the 10 top solutions of each associated translation function for rigid-body refinement cannot be done ‘by hand” as in the previous case, as soon as there is more than one molecule in the asymmetric unit.

In NCS-MR, depending on the number of molecules present in the asymmetric unit, there are thousands of possibilities to be searched. Also, as one is searching with only a fraction of the asymmetric unit, the signal to be expected is intrinsically lower.

AmoRe (23) and other programs like MolRep (34,35) and in fact most Molecular Replacement programs handle this quite effectively in an expert fashion.

MolRep is a very versatile program that has many different options implemented and is part of CCP4 (12).

Obviously, if one wants to try different possible models to increase the chance to find the correct solution, this again makes the search more computer intensive and the best way to deal with this is to follow a given sensible protocol (see below).

As we mentioned already the problem of the rotation function score, leading to a difficult energy landscape to be searched, we can now describe another way to tackle this problem. Since the Translation Function score is much more sensitive, one might try to run a translation for every possible rotation angle, therefore exploring the 6D space exhaustively. The space to be searched in eulerian angles depends on the space group of the crystal and can be found in (36). It turns out that it is doable in most cases within reasonable cpu time with a 'ho rmal" workstation.

There are at least two implementations of this protocol that appeared recently:

-SoMore, which performs a full 6D search with low-resolution data (usually 8 Angstrom), followed by conjugate-gradient minimization of the best solutions (37).

-Systematic rotation+translation searches generated by a script (38) (see also Protocol #6).

Alternatively, one might want to search the 6N-D space (alpha, beta, gamma, tx, ty, tz for each one of the N molecules of the asymmetric unit) using stochastic or *Monte Carlo* methods. In this case, as all molecules are searched simultaneously, the problem of the low signal is less severe than with traditional MR methods. However, the process is quite cpu intensive.

This has been implemented with success by Glykos and Kokkinidis (39-41), who later included a simulated annealing protocol to increase the radius of convergence of the method (Queen of Spades or Qs method).

Other methods have used *genetic algorithms* to search the 6D space: one of them was originated by M. Lewis group (42) and the other resulted in the popular program EPMR (43,44).

The EPMR program is now widely used in the crystallographic community and is especially simple to use (see Protocol #4). This makes it a very attractive candidate for using it in a suite of programs integrated in a web site that goes all the way from model generation to the refinement of the final model (45).

All these methods exploit the fact that there is no need to recalculate the structure factors of the model each time it is rotated or translated: it is in fact sufficient to be able to sample the structure factors at the rotated Miller indices, with or without a phase shift coming from the translation, and this can be done quite effectively by interpolation in reciprocal space (see Protocol #4).

## Protocol #4: Stochastic search methods

### a. Using EPMR (43-44)

```
epmr -m 2 -h 4. -l 12. -n 50 example.cell example.pdb example.hkl >example.log &
```

where example.cell contains the cell parameters and the space group number (19):

```
 70.315    80.852   90.31  90. 90. 90. 19
```

and the model and X-Ray data are in example.pdb and example.hkl, respectively.

The options -l and -h define the low and high resolutions limits of the data, respectively, while the -m option defines the number of molecules to be searched.

The number of different starts is controlled by the -n option.

It is clear from this input lines that epmr is very easy to use and lends itself naturally to be used in a suite of programs aimed at providing quick solution to MR, from model generation to refinement. Indeed, it is the MR package used by B. Rupp's automated protocol (45). It was successful in finding the solution of 6PGL, using default options.

### b. Using Qs (39-41)

```
Qs example.in >example.log &
```

where example.in will look like this:

```
TARGET         CORR-1
CYCLES              10
STEPS          1000000
STARTING_T      0.0150
FINAL_T         0.0050
INFO              1000
NOISE_ADDED       0.10

RESOLUTION         12.0   4.0
AMPLIT_CUTOFF     5.0
SIGMA_CUTOFF      0.0
RANDOM_SELECT   1.0
FREE          0.10

MODEL          example.pdb
DATA           example.hkl
GLOBAL_B       20.0
MOLECULES        2

SEED          357539
SCALECELL       4.0
MAXGRIDSPACING  1.0
SCMODE          wilson
INTERPOLATION   linear

CELL          70.315    80.852   90.31  90. 90. 90.
GROUP          19
```

## How to choose the best model

Even the best possible MR package will fail if the model is not good. Hence, a good deal of efforts should be put into the generation of the best possible model.

The best starting model

Choosing the best model even when only one possible homolog has been detected through alignment and/or threading methods is not an easy task. This is usually dealt with by running MR, in turn, with different versions of the same model.

Up to now, the general accepted rule of thumb was to remove those parts of the model that are believed to be different in the target protein. In other words, it is believed to be better to have an incomplete model with no error than a complete one with errors.

For instance, one could truncate all side chain atoms downstream of the CB (except for glycine) effectively changing the sequence into a poly-Ala; or one could choose to keep the side chain atoms coordinates of only those residues that are strictly conserved between the template and the target and mutate all the others into alanine; or into a serine (changing the CG atom into an OG atom) if there is a conservative substitution.

There are all sorts of possibilities, including the one to keep all atoms inside the core of the molecule and truncating the one with an accessibility to the solvent larger than a given criterion (46).

Truncating loops with high B-factors or non-conserved loops is also possible, although it is only recently that automatic protocols were devised to do the latter using information contained in the multialignement of the sequences (9).

When there are several possible models (templates), the situation becomes more difficult to handle; if there are only 2 or 3, they can be tried individually, say as poly-Ala models.

If there are more, as in NMR structures deposited in the PDB, which usually contain 20 a priori "equivalent" models, specialized protocols have been devised and tested (47), some of which are available on Gerard Kleywegt's web site (http://xray.bmc.uu.se/usf/factory_6.html). One idea is to use an "average" structure; another idea is to weight each atom by a pseudo B-factor which is calculated by an empirical formula that is a function of the rmsd of the position of this atom in the 20 different models of the PDB (see Protocol #5). This usually works well (48).

Protocol #5: How to handle NMR models (47,48)

Goto Gerard Kleywegt's website (Uppsala Software Company):
http://xray.bmc.uu.se/usf/factory_6.html

Get the script in
ftp://xray.bmc.uu.se/pub/gerard/omac/multi_probe

and follow the advices given by Y. W. Chen (47).

### Using homology-modeling derived models

Up to now, models of the protein derived by homology-modeling techniques were not heavily used, because people were reluctant to use models which contain some errors. In these models, all side-chains have been reconstructed, as well as insertions and deletions.

So in a sense, the model is more complete, but it is not certain that this will facilitate the search for the solution of the molecular replacement solution. One fearful feature of homology modeling is that the refinement of the model using standard force-field tends to worsen the model, rather than improve it, at least in test cases. Also, as there are several homology modeling programs available on the web, which all use different methods (distance geometry: Modeller (8), mean-field optimization techniques (49,50)...) that sometimes depend on the generation of random numbers, the question arises as to which should be used.

It is well known that the result of homology modeling is highly dependent on the quality of the alignment between the template and the target. Obviously, the success of MR will be highly dependent on the accuracy of the alignment between the template and the target sequences (51). Hence the need to examine critically the multialignement and, in some cases, to modify it manually.

Luckily, it is now possible to couple the multialignement refining process to the building of models, i.e. process simultaneously 1D and 3D information, and this can be done on the fly. This allows to visualize directly on the 3D level the effect of a modification of the alignment and makes it possible to avoid meaningless alignments (ViTo, 52). However, it demands human intervention and cannot be made automatic.

It should be stressed that homology modeling techniques can handle effectively the case where several structural models (templates) are available.

### Detecting low-homology models (using fold-recognition algorithms)

Recently, as became apparent during the last CASP5 competition (53), methods to detect structural homology with virtually no sequence homology have become more reliable and convincing. They are often based on so-called meta-servers, which address requests to web-based servers of various sorts (secondary prediction methods, threading methods...) and then issue some sort of a consensus score more reliable than any of the separate methods used (10).

One might wonder then if it would be possible to use such methods to pick up remote homologs and use them in MR problems (11). Obviously, the obtained models will contain a lot of errors, so why use them? In particular, homology modeling programs rely heavily on the original backbone coordinates of the template; as there is a well-known exponential law relating the lack of sequence identity between 2 proteins and the rmsd of their coordinates (54), this is worrisome. So at first sight, it might appear that this kind of models would be useless. However, recent reexamination of the same data shows that, if one filters out the outliers in the paired atoms of the structural alignment, the relation between rmsd and lack of sequence identity is no longer exponential but simply linear

(Martin and Labesse, personal communication). Therefore, by truncating judiciously the model, keeping only the conserved core and removing all the variable loops, one might indeed have a useful model.

So altogether, it seems that the trend of conservatively using only models of high sequence identity is now changing, as homology modeling techniques and low-sequence identity structural homologs detection methods are being refined, stimulated by the increase in both the number of sequences and the number of structures, and also because automatic protocols allow for the testing of many different models.

Indeed, given the number of possibilities to generate plausible models, clearly there is room for an automatic protocol trying different things in turn, and then ranking the different solutions. This is precisely what has been done recently in a suite of programs such as the one called CaspR (9), showing very promising results.
This is described in more details in the following section.

## The integrated Molecular Replacement method: comparison of automatic protocols

### CaspR (9)

CaspR is a combination of well-established stand-alone software tools: for a general flowchart of the program, see Fig. 3.
First it reads an MTZ file (CCP4), and extracts from it the unit cell parameters and space group number. Then it runs T-Coffee (55) and 3D-Coffee (56) to get the best possible alignment of the template and the target sequences; in doing so, it identifies variables regions that are likely to be non conserved in the target structure. Then it runs Modeller (8) to construct 10 different models. All these different models are then subjected to AMoRe (23) molecular replacement protocols, with all sorts of different modifications: either truncated from the unreliable regions or not, either as poly Ala or not. Then CNS (25) is used to subject the models with the best MR scores to a round of simulated annealing protocol in internal coordinates space. The status of the procedure can be checked on line at any time. Once submitted, a job gets a reference number that is sent to the user by e-mail and that has to be quoted for consulting the job status. The data are erased after one week.
It is our experience that even if the solution can be obtained with more traditional MR packages, the final R-factor and R-free obtained with CaspR are usually much lower, thereby effectively reducing the time spent in manual reconstruction in front of the graphics display terminal.
For instance, in the 6PGL case, the best score of the MR itself was obtained with the truncated Model #2 of Modeller (Corr. Coeff.=38%), but the best refined model was Model #23 with Rwork=43.8% and Rfree=52%; this kind of score could never be obtained with only Rigid Body refinement.

TB Consortium (45)

In the case of the TB Structural Genomics Consortium (45), PSI-BLAST is used to find homolog(s) in the PDB, CLUSTAL-W is used for obtaining a multialignment, AL2TS to build the model(s), EPMR (43,44) to do Molecular Replacement and CNS (25) to refine the model(s). In addition, the authors use their own local version of ARP/wARP, called Shake-and-wARP, to attempt automatic building of the model in the electron density map.

Clearly, many combinations of such protocols are possible, by using different software at each step of the whole process.

G. Labesse and L. Martin (GTBio, Lyon, June 2004)

In this case, the @TOME meta server (10) is used to detect low structural homology three-dimensional models, as a consensus between six different “threading servers”. Then, after model building by Modeller (8), MolRep (34,35) is run, followed by refinement using Refmac (2).

In some instances, this suite of programs was able to find the MR solution using a model with sequence identity as low as 20% (G. Labesse, personal communication).

## Making the most of your model using Normal Modes: Pushing towards the systematic exploration of structural diversity

In addition, a new way to take into account the structural variability of protein structures is being actively explored, with very promising results.

It is well known that many proteins can exist in several structural states, such as open or closed, depending on the presence of one of their substrates. The rmsd between the two forms is such that if one tries to solve the MR with the wrong form, MR will fail, while it might succeed if one uses the other one. Of course, one might try to use both, but it may well be that all accessible structural states of a given protein are not deposited in the PDB.

Hinge-bending as well as shear movements have been documented over the years, with rmsd between the two forms of the same proteins spanning a very large range. While hinge-bending movements can in principle be tackled by dividing the protein into individual domains and then solving the MR problem for each of the domains, other types of movements are very difficult to deal with, especially as they are very collective and involve the coordinated movement of many atoms. Recently, however, it has become apparent that most structural transitions observed in the PDB can be modeled quite accurately by the low-frequency Normal Modes derived from a simplified representation of the protein, namely the Elastic Network Model (57). The Elastic Network Model (Fig. 4) is a simplified representation of proteins where each residue is represented by a single point that is linked to its spatial neighbors (within a given radius usually taken as 8-12

Angstroms) by a spring of constant strength (58); it works surprisingly well despite its simplicity and is by construction most apt to model collective movements (59). In fact, it has been shown that most movements can be modeled with on the average only 2 normal modes (57). As a result of advances in the computation of such modes, there is virtually no limit in the size of the macromolecular assembly for which low-frequency normal modes can be calculated (60).

Recently, two independent applications of these ideas to MR have been described (Fig. 5):

-one is intended to produce structural diversity in the model by grid-sampling the amplitudes of (at most) two of the lowest-frequency normal modes; all the produced models are subjected to a standard MR protocol, leading to solutions in cases where all the other methods failed; this is now available on line as a web server called elNemo (61,62).

-the other is a refinement program using a correlation coefficient with the experimental data as a target and conjugate-gradient algorithm as the engine, while the only degrees of freedom are the amplitudes of the lowest frequency Normal Modes (26). It is intended to be used as a generalization of Rigid Body refinement, as the 6 degrees of freedom of global rotation and translation are also included as Normal Modes of null frequency. Convincing test cases as well as real cases have shown that it works well and, although it was originally described to work with C-alpha-only models, it has now been modified to include all atoms in the refinement and has also been put on line as a web-server (http://lorentz.immstr.pasteur.fr).

**How to get the best (least-biased) starting phases**

Once the best possible molecular replacement solution has been found and refined by e.g. CNS simulated annealing protocol in internal coordinates space, there is a need to use the least possible biased phases, as recently emphasized in (45).

This is usually done by using SIGMAA-weighted maps, originally due to R. Read (63,64) and implemented in most crystallographic packages (Refmac5 (2), CNS (25)...).

Then, one can go on and try to use directly automatic construction methods such as ARP/wARP (3) or Resolve (4).

However, we wish to point out here that there exists a very general formalism based on mean-field theory and statistical mechanics (65) which in principle should allow to get rid of most of the errors contained in the model phases and which points to the use of other kinds of weighted coefficients, calculated in a self-consistent manner, as inputs to the Fourier transforms to calculate the map. Even though the method still remains to be implemented in a space-group general program, preliminary results in the space group P2 (1)2(1)2(1) show great promise (Delarue, unpublished work).

**Special Cases: phased TF and locked RF.**

If, for some reason, molecular replacement failed, leaving no option but to look for heavy atom derivatives and if, for lack of isomorphism, the phasing power of these derivatives do not allow for a straightforward interpretation of the map, it is still possible to use the experimental phases in conjunction with the model at hand. Indeed, it is in general much easier to place a model in even a poor  experimental map than without phases. This can be done in reciprocal space using an analog of the translation function, called the phased translation function (66). This is simply a correlation coefficient between maps of the model placed at different origins and the experimental one. This can be carried out in reciprocal space, where all the products of type ($\mathbf{F}_{obs}.\mathbf{F}_{model}$) have been replaced by ($\mathbf{F}_{obs}.\mathbf{F}_{model}^{*}$), where $\mathbf{F}_{model}^{*}$ is the complex conjugate of $\mathbf{F}_{model}$. Once the model has been placed correctly in the unit cell of the crystal, experimental phases can be combined with phases from the model, followed by solvent flattening using a molecular envelope derived directly from the model. This usually results in a much better map. However, the rotation function is still in general no easier to solve than without the experimental phases, so one might think that all of this is of little use. There is however a way out of it, which is to scan the entire rotation function space with the phased translation search as a score (see Protocol #6). The present author used it in a number of cases with success: it is in general computationally doable and leads to a very clear signal (67,68).

Protocol #6: Full 6D search with the phased translation function.

1. Determine the asymmetric unit of the space group of the rotation function of your space group (31).

2. Write and execute jiffy code to create the (formatted) input file to explore exhaustively rotation space:

```
c...
      delta0=5.
      delta1=delta0/2.
      do i=-imax,imax
        do j=0,jmax
          do k=-kmax,kmax
               alpha  =alpha0     + delta0*i
               beta    = beta0      + delta1*j
               gamma=gamma0 + delta0*k
               write(2,'  (3f10.5,4x,2f11.3,i10)')alpha,beta,gamma,r,t,iu
          enddo
        enddo
      enddo
```

c...
The resulting formatted file will be fed to AMoRe as m1.rts, the input file to Phased Translation Function (PTF).

3. Run AMoRe PTF in all the possible space groups (see Protocol #1), after having run a test case (as in Protocol #2).

4. Use perl to extract the information on the solutions with the highest score from the (enormous) log file. Sort the solutions.

5. Plot results using gnuplot.

---

Another special case where the signal of the rotation function could be enhanced concerns crystals where the self-rotation function can be interpreted without ambiguity; in this case, the so-called locked cross-rotation function (69) allows to search for cross rotations which are compatible with the self rotation function. This usually results in a much better signal-to-noise ratio.

## Concluding Remarks

It seems pretty clear that easy-to-use and web-interfaced automated protocols will be more and more useful in the next future. The possibility of combining different tools at different steps of the integrated process makes it inevitable that more of them will be developed and open to the public. What' sgood about this kind of approach is that the authors can keep track of both the successes and failures of jobs submitted by the crystallographic community and use these as bench marks and opportunities to improve their protocols. So their performances should keep growing, provided that crystallographers don't shy away from them because of confidentiality problems...

Finally, by way of setting perspectives, we wish to point out that the only alternative to the use of many different models successively, with the same protocol, until one of them gives a better signal, would be to use all of them simultaneously, weighted by some linear prefactor that remains to be determined. The sum of all these prefactors for all the different copies of the model should of course be 1. This is actually reminiscent of some recent approach in the refinement of macromolecules using a multi-copy strategy, so as to take into account some inherent flexibility of the model (70,71). We suggest that refining the weights of the different models at each orientation of the model in the rotation function should improve the signal-to-noise ratio of the whole procedure. Translation function searches should then proceed normally using the best combination of weighted model(s) identified in this way.

## Acknowledgments

We gratefully acknowledge financial support from ACI-IMPB045 from the Ministere de la Recherche et de la Technologie and from GDR 2417 du C.N.R.S.
We thank G. Labesse (CBS, CNRS, Montpellier) for discussions and communication of results prior to publication and P. Dumas (IBMC, CNRS, Strasbourg) for comments on the manuscript.

**Legend of Figures**

Figure 1
Histogram of the number of articles in Acta Cryst. D containing “Molecular Replacement” in it s title or abstract, year by year. The score in 2004 is a projection.

Figure 2
Definition of the Molecular Replacement problem and the six degrees of freedom needed to describe it.

Figure 3
General flow chart for an integrated Molecular Replacement structure solution protocol.

Figure 4
Illustration of the Elastic Network Model for TMP kinase (cutoff=8 Angstroms).

Figure 5
Grid sampling and refinement of the amplitudes along, successively, the first ten Normal Modes for the Glutamine-binding protein. The open structure (1WDN) is used to calculate Normal Modes and then deformed along these Normal Modes to fit X-Ray data of the closed form (1GGG PDB code). The amplitudes are scanned from -500 to 500, using 100 grid points for each normal mode. Once the best amplitude is found for a given Normal Mode, it remains fixed for the following searches along the next available Normal Modes. It is apparent from this plot that no further increase of the correlation coefficient is obtained after the 5th mode has been added. The rmsd between 1WDN and 1GGG is 5.5 Angstroms. After finding the amplitudes of the 10 lowest frequency Normal Mode using this grid-based approach, Conjugate Gradient refinement is used to further improve the correlation coefficient, reaching ultimately 26%. The initial score was 6%.
The same example is also treated in (61), with similar results.

Table 1
List of web sites for MR packages with a short description of their distinctive features.

## Bibliography

1. Jones, T.A., Zhou, J.Y., Cowan, S.W., and Kjelgaard, M. (1991). *Acta Cryst.*, **A47**, 110.
2. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Acta Cryst., D53, 240.
3. Perrakis, A., Harkiolaki, M., Wilson, K.S., and Lamzin, V.S. (2001). *Acta Cryst.*, **D57**, 1445.
4. Terwilliger, T. (2003). *Acta Cryst.*, **D59**, 1174 and *ibid*, 39 and *ibid*, 45.
5. Adams, P., Pannu, N.S., Read, R.J., and Brunger, A.T. (1999). *Acta Cryst.*, **D55**, 181.
6. 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**, 235-242.
7. Altschul S.F, Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., and Lipman D.J. (1997). *Nucleic Acids Res.*, **25**, 3389.
8. Sali, A. and Blundell, T. L. (1993). *J. Mol. Biol.,* **234**, 779-815.
9. Claude J.B., Suhre K., Notredame C., Claverie J.M., and Abergel C. *(2004). Nucleic Acids Res.,* **32,** W606.
10. Douguet, M. and Labesse, G. (2001). *Bioinformatics*, 17, 752.
11. Jones, D.T. (2001). *Acta Cryst.*, **D57**, 1428.
12. Collaborative Computational Project. (1994). *Acta Cryst.*, **D50**, 760.
13. Lee, S., Sawaya, M.R., and Eisenberg, D. (2003). *Acta Cryst.*, **D59**, 2191.
14. Rossmann, M.G. (1990). *Acta Cryst.*, **A46**, 73.
15. Rossmann, M.G. and Blow, D.M. (1962). *Acta Cryst.*,**15**, 24.
16. Crowther, R.A. (1972). in *The molecular replacement method.* Edited by M. G. Rossmann. p. 173. New York: Gordon and Breach.
17. Navaza, J. (2001). in *International Tables of Crystallography*, vol. F, p. 269.
18. Crowther, R.A. and Blow, D.M. (1967). *Acta Cryst.*, **23**, 544.
19. Huber, R. and Schneider. (1985). *J. Appl. Cryst.*, **18**, 165.
20. Navaza, J. and Saludjian, P. (1997). in *Methods in Enzymology*, vol 276, p. 581.
21. Fokine, A. and Urzhumtsev, A. (2002). *Acta Cryst.*, **D58**, 72.
22. Urzhumtseva, L.M. and Urzhumtsev, A.G. (1997). *J. Appl. Cryst.*, **30**, 402.
23. Navaza, J. (2001). *Acta Cryst.*, **D57**, 1367.
24. Jamrog, D.C., Zhang, Y., and Phillips Jr., G.N. (2004). *Acta Cryst.*, A**60**, 214.
25. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T., and Warren G.L. (1998). *Acta Cryst.*, **D54**, 905-921.
26. Delarue, M. and Dumas, P. (2004). *Proc. Natl. Acad. Sci. (USA),* **101**, 6957.
27. Harada, Y., Lifchitz, A., Berthou, J., and Jolles, P. *(1981). Acta Cryst.,* A**37**, 398.
28. Stubbs, M.T. and Huber, R. (1991). *Acta Cryst.*, **A47**, 521.
29. Brunger, A.T. (1997). in *Methods in Enzymology*, vol 277, p. 366.
30. DeLano, W.L. and Brunger, A.T. (1995). *Acta Cryst.*, **D51**, 740.

31. Urzhumtsev, A. and Urzhumtseva, L. (2002). *Acta Cryst.*, **D58**, 2066.
32. Read, R.J. (2001). *Acta Cryst.*, **D57**, 1373.
33. Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). *Acta Cryst.*, **D60**, 432.
34. Vagin, A. and Teplyakov, A. (1997). *J. Appl. Cryst.*, **30**, 1022.
35. Vagin, A. and Teplyakov, A. (2000). *Acta Cryst.*, **D56**, 1622.
36. Rao, S.N., Jin, J.H., and Hartsuck, J.A. (1980). *Acta Cryst.*, **A36**, 878.
37. Jamrog, D.C., Zhang, Y., and Phillips Jr, G.N. (2003). *Acta Cryst.*, **D59**, 304.
38. Sheriff, S., Klei, H.E., and Davis, M.E. (1999). *J. Appl. Cryst.*, **32**, 98.
39. Glykos, N.M. and Kokkinidis, M. (2003). Acta *Cryst.*, **D59**, 709.
40. Glykos, N.M. and Kokkinidis, M. (2001). *Acta Cryst.*, **D57**, 1462.
41. Glykos, N.M. and Kokkinidis, M. (2000). *Acta Cryst.*, **D56**, 169.
42. Chang, G. and Lewis, M. (1997). *Acta Cryst.*, **D53**, 279.
43. Kissinger, C.R., Gelhaar, D.K., and Fogel, D.B. (1999). *Acta Cryst.*,**D55**, 484.
44. Kissinger, C.R., Gelhaar, D.K., Smith, B.A., and Fogel, D.B. (2001). *Acta Cryst.*, **D57**, 1474.
45. Rupp, B. et al., (2002). *Acta Cryst.*, **D58**, 1514. http://www.doe-mbi.ucla.edu/TB/
46. Delarue, M., Samama, J.P., Mourey, L., and Moras, D. (1990). *Acta Cryst.*, **B46**, 550.
47. Chen, Y.W. (2001). *Acta Cryst.*, **D57**, 1457.
48. Wilmanns, M. and Nilges, M. (1999). *Acta Cryst.*, **D52**, 973.
49. Koehl, P. and Delarue, M. (1995). *Nature Struct. Biol.*, **2**, 163;
50. Koehl, P. and Delarue, M. (1994). *J. Mol. Biol.*, **239**, 249.
51. Schwarzenbacher, R., Godzik, A., Grzechnik, S.K., and Jaroszewski, L. (2004). *Acta Cryst.*, **D60**, 1229.
52. Catherinot, V. and Labesse, G. (2004). *Bioinformatics*, in press.
53. CASP5. (2003). *Proteins*, **53**, Suppl 6.
54. Chothia, C. and Lesk, A. (1986). *EMBO J.* , **5**, 823.
55. Notredame, C., Higgins, D., and Heringa, J. (2000). *J. Mol. Biol.*, **302**, 205-217.
56. O' Sullvan, O., Suhre, K., Abergel, C., Higgins, D.G., and Notredame, C. (2004). *J. Mol. Biol.*, **340**, 385.
57. Krebs, W.G., Alexandrov, V., Wilson, C.A., Echols, N, Yu, H., and Gerstein, M. (2002). *Proteins*, **48**, 682.
58. Tirion, M. (1996). *Phys. Rev. Lett.*, **77**, 1905.
59. Delarue, M. and Sanejouand, Y.H. (2002). *J. Mol. Biol.*, **320**, 101.
60. Tama, F. and Sanejouand, Y.H. (2001). *Protein Engng.*, **14**, 1.
61. Suhre, K. and Sanejouand, Y.H. (2004). *Acta Cryst.*, **D60**, 796.
62. Suhre, K. and Sanejouand, Y.H. (2004). *Nucleic Acids Res.*, **32**, W606.
63. Read, R.J. (1986). *Acta Cryst.*, **A42**, 140.
64. Read, R.J. (1990). *Acta Cryst.*, **A46**, 900.
65. Delarue, M. and Orland, H. (2000). *Acta Cryst.*, **A56**, 562.

66. Bentley, G. (1997). *Methods in Enzymology*, **276**, 611.

67. Li de la Sierra, Munier-Lehmann, H., Gilles, A.M., Barzu, O., and Delarue, M. (2001). *J. Mol. Biol.*, **311**, 87.

68. Delarue, M., Boule, J.B., Lescar, J. Expert-Bezancon, N., Jourdan, N., Sukumar, N., Rougeon, F., and Papanicolaou, C. (2002). *EMBO J.*, **21**, 427.

69. Tong, L. and Rossmann, M.G. (1997). in *Methods in Enzymology*, vol 276, p. 594.

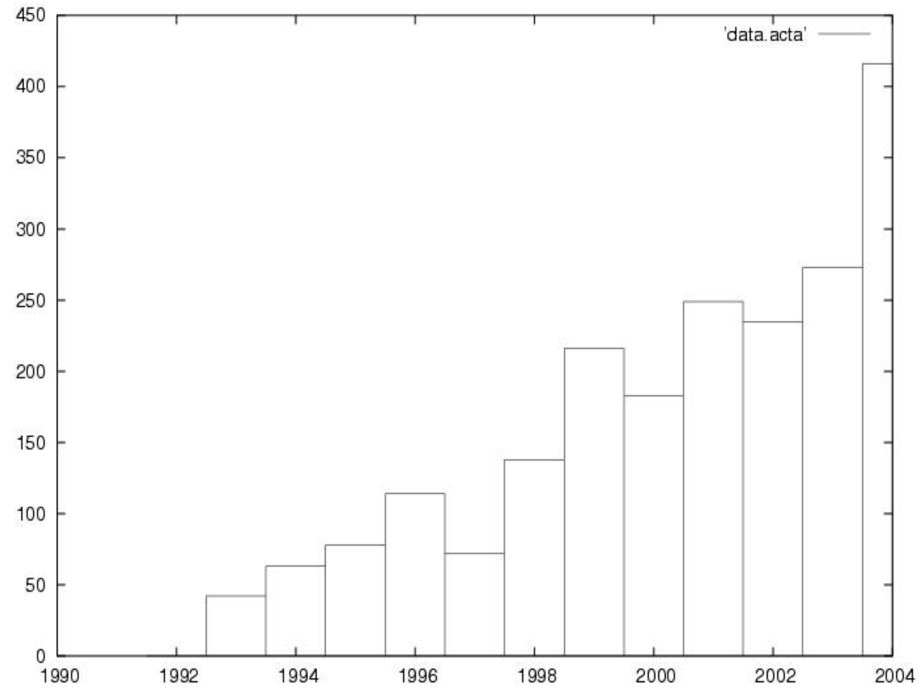
70. Burling, F.T. and Brunger, A.T. (1994). *Isr. J. Chem.*, **34**, 165.

71. Pellegrini, M., Gronbech-Jensen, N., Kelly, J.A., Pfluegl, G., and Yeates, T.O. (1997). *Proteins*, **29**, 426.

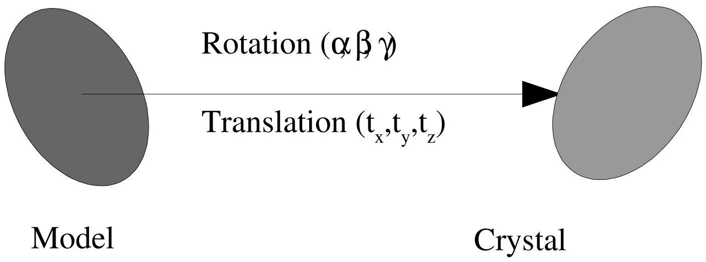
| ***Package Name*** | ***Search space*** | ***NCS*** | ***Web site or E-mail of the author*** | ***Model(s) generation*** | ***Post-ref.*** |
|---|---|---|---|---|---|
| AMoRe | 2x3D | Yes | jorge.navaza@gv.cnrs-gif.fr | No | RigBod |
| CNS | 2x3D | Yes | http://cns.csb.yale.edu/v1.1/ | No | CNS |
| MolRep | 2x3D | Yes | http://www.ysbl.york.ac.uk/~alexei/molrep.html | Some | RigBod |
| EPMR | 6D | Yes | http://www.msg.ucsf.edu/local/programs/epmr/epmr.html | No | Conj. Grad. |
| Qs | 6D | Yes | http://origin.imbb.forth.gr:8888/~glykos/Qs.html | No | Sim. Anneal |
| SoMore | 6D | No | http://www.caam.rice.edu/~djamrog/somore.html | No | BFGS |
| Phaser | 2x3D | Yes | http://www-structmed.cimr.cam.ac.uk/phaser | No | No |
| CaspR | 2x3D | Yes | http://igs-server.cnrs-mrs.fr/Caspr/index.cgi | Yes | CNS |
| @TOME | 2x3D | Yes | http://bioserv.cbs.cnrs.fr/HTML_BIO/frame_meta.html | Yes | No |

**Table 1**

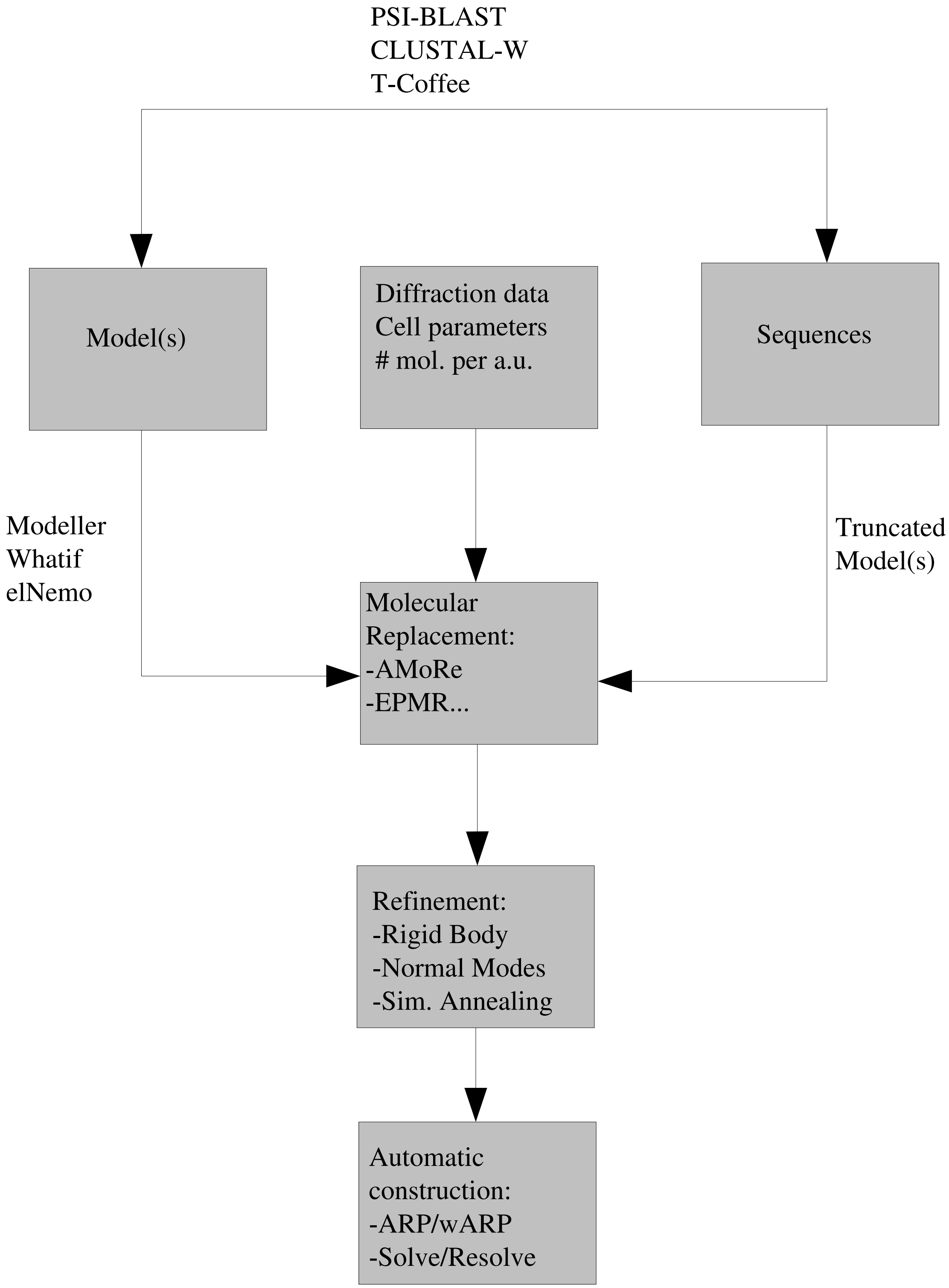
## Figure 1


'data.acta'
450
400
350
300
250
200
150
100
50
0
1990
1992
1994
1996
1998
2000
2002
2004

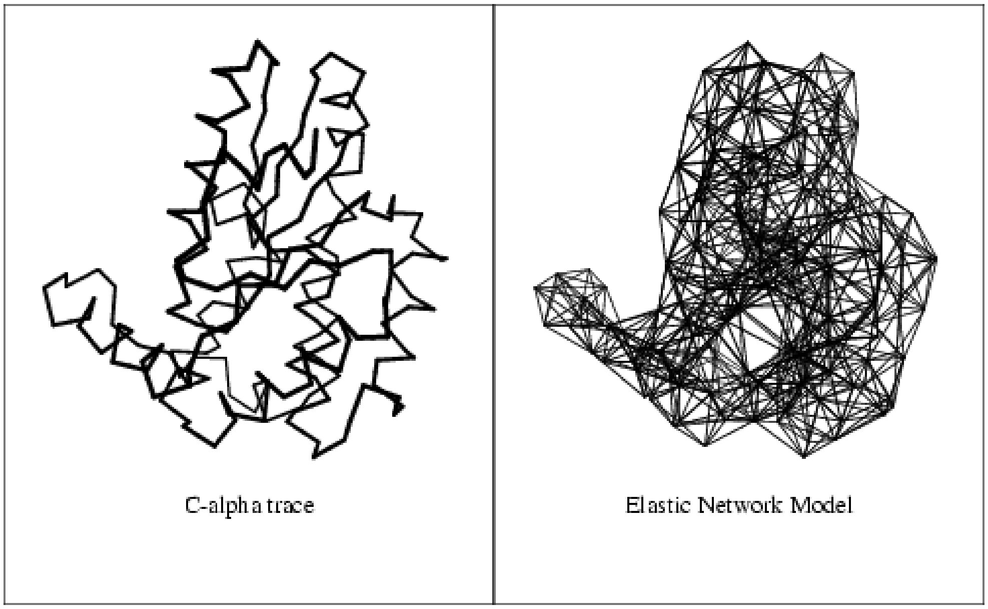
# Figure 2



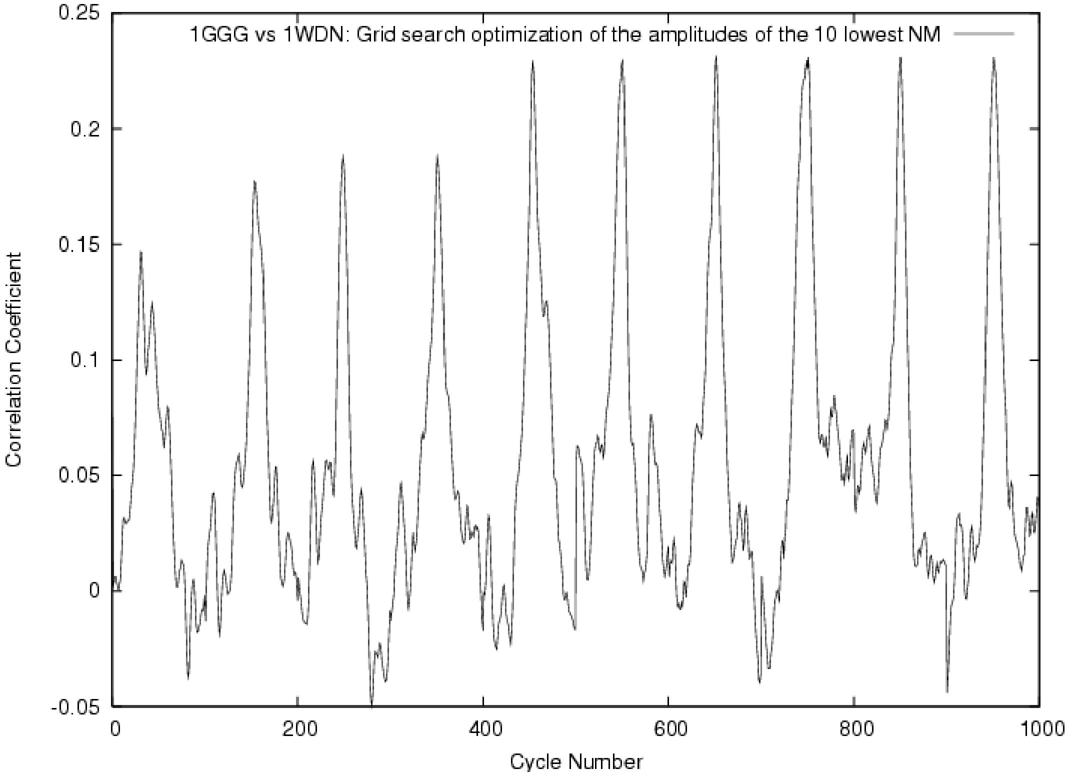
## Figure 3


PSI-BLAST
CLUSTAL-W
T-Coffee
Model(s)
Diffraction data
Cell parameters
# mol. per a.u.
Sequences
Modeller
Whatif
elNemo
Truncated
Model(s)
Molecular
Replacement:
-AMoRe
-EPMR...
Refinement:
-Rigid Body
-Normal Modes
-Sim. Annealing
Automatic
construction:
-ARP/wARP
-Solve/Resolve

## Figure 4



# Figure 5


1GGG vs 1WDN: Grid search optimization of the amplitudes of the 10 lowest NM
Correlation Coefficient
0.25
0.2
0.15
0.1
0.05
0
-0.05
0
200
400
600
800
1000
Cycle Number