Equation Chapter 1 Section 1 *Chris Oostenbrink, January 7, 2004*

NOE

A short overview of NOE analysis from MD simulations will be given. From the experimental measurements to the NOE violations or distance restraints, a lot of aspects, assumptions and corrections come together. This is a summary of some literature, notes by Wilfred, discussions with Lorna Smith and the GROMOS96 manual. Although the description should be generally applicable, some examples will be taken from a set of Lysozyme NOE's.[1,2]

Experiment

From the peak heights of cross peaks in multidimensional NMR experiments, one can estimate NOE intensities for the interaction between two spins I and S. In many cases, these are then divided into a number of categories (strong, medium, weak or very weak).[1] If the cross peaks have been assigned to specific hydrogen atoms, one can then define lower and upper bounds to the distance between these hydrogen atoms. The following table lists the upper and lower bounds for the different categories of NOE intensities as they have been applied for the lysozyme.[2] The lower bound is always set to 1.8 Å, as would correspond to a van der Waals contact between to spins.

NOE intensity	lower bound (Å)	upper bound (Å)
strong	1.8	2.5
medium	1.8	3.0
weak	1.8	4.5
very weak	1.8	5.5
even weaker*	1.8	7.5

* cross peaks that were only observed in 1H NOESY spectra with a long 500 ms mixing time

Average distances

The upper and lower bounds to the average interatomic distances can be used in molecular simulations in two ways. First, one can use the bounds as a tool for structure determinations by running a simulation while applying restraints on the interatomic distances. Second, one can perform a free simulation and check whether the interatomic distances on average satisfy these restraints, as a force field test. This would be an NOE analysis of an MD trajectory.

The experimental cross peak intensity and resulting NOE distances, represent an average over space and time. Often, one sees the definition of an effective distance as sensed by the NOE, due to the interaction between spins I and S

$$r_{IS}^{eff} = \left\langle r_{IS}^{-p} \right\rangle^{-1/p} \tag{1}$$

Care has to be taken with respect to the averaging exponent p of the interatomic distances.[3] The NOE intensity is proportional to the spectral density at zero frequency

$$J(0) = \int_0^\infty C(t)dt \tag{2}$$

where C(t) is the time correlation function decribing the orientation $\theta(t)$ and length $r(t)$ of the interatomic vector connecting the two nuclei:

$$C(t) = \left\langle \frac{\cos^2\theta(t') - 1}{2r^3(t')} \frac{\cos^2\theta(t'+t) - 1}{2r^3(t'+t)} \right\rangle_{t'} \tag{3}$$

Generally, $C(t)$ has a fast component due to fast fluctuations and one or more slower components, e.g. due to overall rotation of the molecule. Now we assume that the fast component is due to fluctuations in $r(t)$ and the slow components are due to fluctuations in $\theta(t)$ and that these are not coupled. In that case we factorize the correlation function

$$C(t) = C_r(t) \cdot C_\theta(t) \tag{4}$$

with

$$\begin{aligned} C_r(t) &= \left\langle r^{-3}(t')r^{-3}(t'+t) \right\rangle_{t'} \\ C_\theta(t) &= \left\langle \tfrac{1}{2}(3\cos^2\theta(t') - 1)\tfrac{1}{2}(3\cos^2\theta(t'+t) - 1) \right\rangle_{t'} \end{aligned} \tag{5}$$

and we can write

$$\lim_{t\to 0} C_r(t) = \left\langle r^{-6} \right\rangle$$
$$\lim_{t\to \infty} C_r(t) = \left\langle r^{-3} \right\rangle^2 \tag{6}$$

If we assume that the fast part of $C(t)$ decays exponentially with relaxation time $\tau_{fast} = \tau_r$,

$$C_r(t) = \left[\left\langle r^{-6} \right\rangle - \left\langle r^{-3} \right\rangle^2 \right] e^{-t/\tau_r} + \left\langle r^{-3} \right\rangle^2 \tag{7}$$

and similarly that the slow part decays exponentially with $\tau_{slow} = \tau_\theta$,

$$C_\theta(t) = \left\langle \left[\tfrac{1}{2}(3\cos^2\theta - 1) \right]^2 \right\rangle e^{-t/\tau_\theta} \tag{8}$$

the integral (2) can be carried out ($\tau_r << \tau_\theta$):

$$J(0) = \left[\left\langle r^{-6} \right\rangle - \left\langle r^{-3} \right\rangle^2 \right] \tau_r + \left\langle r^{-3} \right\rangle^2 \left\langle \left[\tfrac{1}{2}(3\cos^2\theta - 1) \right]^2 \right\rangle \tau_\theta \tag{9}$$

Since τ_r is very small, the first term in (9) can be neglected, and the NOE becomes proportional to $\left\langle r^{-3} \right\rangle^2$.

This is why in MD simulations, which generally cover only the fast distance vibrations, the averaging, also in time-averaging refinement, is performed as $\left\langle r^{-3} \right\rangle^{-1/3}$ and not as $\left\langle r^{-6} \right\rangle^{-1/6}$.

When do the mentioned assumptions break down?

1. For small molecules, when the molecular tumbling is faster than for proteins
2. When angular motions are fast, *e.g.* for hydrogen atoms
3. When there is coupling between the distance and angular motions, *e.g.* NOEs between atoms of differently rotating side chains
4. When the distance motions is as slow a the other ones, *e.g.* for interchanging distances

In the Fletcher paper[4], however, it is written: “Although many internal motions may be significantly faster than overall tumbling, methyl group rotations are the only common example where a group of spins *are made equivalent* as a result of such internal motion…” Maybe this reasoning should make us reconsider?

Corrections: pseudo atoms and multiplicity

So far, everything seems to be quite straightforward, unfortunately both the experiment and the simulation poses some problems, that require ones attention and possibly corrections to be added to the experimental upper bounds.

From the experimental side, one can often not distinguish individual hydrogen atoms, either because they are genuinely indistinguishable (*e.g.* the protons in a methyl group), or because one cannot assign the stereospecific hydrogen atoms (*e.g.* the β-hydrogens in an amino acid side chain). Which distance should one calculate if the experiment cannot tell from which hydrogen the NOE signal comes? We generally take the 'centre average' approach[5]: a "constraint for several protons is referred to a single *pseudo atom* at the mean position of the atoms in the equivalent group."[4] This can give rise to pseudo atom corrections and multiplicity corrections. Alternative approaches such as 'r^{-6} averaging' or 'r^{-6} summing' have also been reported, but are beyond the scope of this note.[4]

In the centre averaging approach, the hydrogen atoms of an equivalent group I are represented using a pseudo atom Q at the mean position of the individual atoms, I_i. If we consider the NOE due to the interactions of these atoms with another hydrogen atom S, we can easily see, that the distance r_{QS} between the pseudo atom and S is almost always larger than the shortest of the individual distances r_{IiS}. But this shortest distance is actually contributing most to the NOE signal, from which the upper bound is calculated. This means that we have to correct the upper bound for the fact that we are calculating the distance to Q rather than the closest I_i. Originally, the pseudo atom corrections were defined as the distance between any one of the equivalent atoms I_i and the pseudo atom Q.[5] This is always sufficiently long, because it is the longest distance that the pseudo atom could be further away from S than the closest of the I_i. However, in ref [4] it is argued that this definition correspond to the assumption that *only* the closest of the individual spins I_i contributes to the NOE interaction with S. In this paper another definition is suggested as the maximum possible value of $r_{QS} - r^{eff}$, thus taking all members of the equivalent group into account. The correction is now the maximum distance by which the distance from S to Q can exceed the averaged distance 'sensed' by the NOE and used to set the upper bound. The difficulty with such a definition is that r^{eff} as calculated from eq.

(1) depends on the geometry of the group, making the determination of the pseudo atom correction not always straightforward.[4]

A multiplicity correction is applied when an NOE is known to come from an average signal over several hydrogen atoms. For an NOE between two groups of hydrogen atoms I and S, the upper bound is usually scaled by a factor

$$Z = (n_I n_S)^{1/p} \tag{10}$$

where p corresponds to the averaging power (3 or 6), n_I and n_S are the number of equivalent hydrogen atoms in each group. E.g. n_I is 3 for a methyl group, 6 for a rotating iso-propyl group or 2 for a flipping atomatic ring. Instead of this multiplication of the upper bound by Z, sometimes another distance is added to it[6] or it is ignored altogether.

Using both pseudo atom corrections and multiplicity corrections it is easy to stretch any NOE distance to long distances. In ref [4], the authors argue that using their new definitions of the pseudo atom corrections together with a multiplicity correction will usually result in a tighter corrected upper bound than the original pseudo atom correction for a methyl group if the uncorrected upper bound is less than 5.5 Å. If both types of corrections are applied, the uncorrected upper bound is *first* multiplied by Z in eq. (10) and *then* the pseudo atom correction is added.

MD simulations: virtual atoms

An additional problem for MD simulations with a united atom force field is that many of the hydrogen atoms for which the NOE distances are measured are not simulated explicitly. To overcome this problem, one can construct the position of an implicit hydrogen atom from the geometry of the explicit heavy atoms, creating a *virtual atom*. The calculation of the position of virtual atoms, can require the positions of up to 4 heavy atoms. In the case of structure refinement based on NOE distance restraints, virtual atoms impose the additional difficulty that forces calculated on the virtual atom need to be redistributed to the heavy atoms that define its position. See the GROMOS96 manual for further details.[7]

Combining the different kinds of pseudo and virtual atoms, a number of types have been defined in GROMOS, which should be chosen according to experimental assignment of the NOE intensity.

0. Explicit hydrogens. If the experimental NOE describes a single hydrogen atom that is explicitly simulated (polar or aromatic hydrogens), no virtual or pseudo atoms play a role and the position of the hydrogen atom can be taken directly for the calculation of the NOE distance.
1. CH1 (aliphatic). A single hydrogen that can be calculated as a virtual atom. There will be no need for pseudo atom or multiplicity corrections.
2. CH1 (aromatic). A single virtual hydrogen atom bound to an *sp*2 hybridized carbon atom. In the current GROMOS force fields, these hydrogen atoms are usually treated as explicit atoms, making this type obsolete. No pseudo atom or multiplicity corrections are needed.
3. CH2 (non-stereospecific). Two aliphatic hydrogens are treated as a single pseudo atom. In ref [4] the authors distinguish three cases:
   a. Both atoms are resolved and measured, but have the same intensity. In this case the average is also the same, so one should not use a multiplicity correction, but only add a pseudo atom correction.
   b. Both atoms are resolved and measured, with different intensities, giving rise to two upper bounds A and B (A<B). A pseudo atom approach is suggested, using an upper bound of $2^{1/p}\times$A *or* just B, whichever is shorter. A pseudo atom correction is also needed.
   c. One peak is resolved, giving only one signal, which is either an average or a single peak. A multiplicity correction of $2^{1/p}$ is suggested as well as a pseudo atom correction.

   For a simulator without direct access to the experimental spectra, it will be very difficult to determine which of these cases is relevant, if the only assignment we have is *e.g.* 'HB*'. My suggestion would be to apply option c as it is the most straightforward and it will lead to the highest upper bound, thus effectively ignoring data that we are not sure of.

4. CH2 (stereospecific). Two aliphatic virtual hydrogen atoms. Again ref [4] distinguishes two cases
    a. If there are two different upper bounds specified, the authors suggest to use only the stronger interaction (shorter distance) to define an upper bound because of the risk of spin diffusion. Unfortunately, we will still have trouble automatically assigning the correct virtual atom to the correct signal. Currently the prepnoe program (see below) generates two distances to monitor, which can then later be assigned and reduced to a single distance.
    b. If the two protons have the same upper bounds, the authors suggest to treat them as a type 3a, i.e. as a pseudo atom with pseudo atom correction, but no multiplicity correction.
5. CH3. Three genuinely indistinguishable protons, described by a single pseudo atom. Both a multiplicity correction and a pseudo atom correction are required.
6. iso-propyl. In case of a fastly rotating iso-propyl group, the six protons are indistinguishable and are treated as a single pseudo atom. Both a multiplicity correction and a pseudo atom correction are applied.
7. flipping aromatic ring. Even though the aromatic hydrogen atoms are usually explicitly simulated, the experiment can not distinguish between them. We therefore treat them as a pseudo atom, which is located at the aromatic carbon in between the hydrogens. Both a multiplicity correction and a pseudo atom correction are applied.

In addition, we might want to consider to define another type

8. NH2 (planar). From heteronuclear NMR experiments, involving N-H spin coupling, the two hydrogen atoms of e.g. Asn and Gln are often seen, but not distinguished. In these cases we should treat these explicit hydrogen atoms as a pseudo atom and add a multiplicity correction and a pseudo atom correction.

Additional types that can be thought of[4], have so far not been seen in experimental data that was analysed by us.

For each of the above mentioned types, the program prepnoe (see below) offers the possibility to apply corrections. In the following table some corrections that have been found in the literature are listed

Corrections to experimental upper bounds. All distances in Å.

	Wüthrich[5]		Fletcher[4]		GROMOS96[7]		Koning[8]	
Type	pseudo	multipl.	psuedo	multipl.	pseudo	multipl.	pseudo	multipl.
0								
1								
2								
3a	1.0		0.7	1	0.9			
3b			0.7	$2^{1/p}/1^{\dagger}$				
3c			0.7	$2^{1/p}$				
4a								
4b			0.7	1				
5	1.0	+0.5 *	0.4	$3^{1/p}$	1.0		0.3	$3^{1/p}$
6	2.4	+0.5 *	1.5	$6^{1/p}$	2.2			
7	2.0		2.0	$2^{1/p}$	2.1			
8	1.0		0.7	$2^{1/p}$				

* No multiplication, but an additional correction of 0.5 Å was added.[6]

† For two signals *A* and *B* with different intensity (*A*<*B*) take MIN($2^{1/p}A$, *B*)

Note that all pseudo atom corrections have been obtained by assuming bond lengths and configurations that do not necessarily correspond to the GROMOS bond lengths and bond angles. This means that simply taking alternative correction factors might be a dangerous undertaking.

Practical

In this section I will try to give an outline of an NOE analysis starting from the experimental data, using gromos++ programs. As an example I have taken the Lysozyme NOE's.[2]

The experimental data often comes in the XPLOR format, which looks something like

```
ASSIGN (RESID   1    AND NAME HA   ) (RESID   2    AND NAME HN   )   2.2  0.4  0.3
ASSIGN (RESID   2    AND NAME HN   ) (RESID   2    AND NAME HA   )   4.2  2.4  0.3
ASSIGN (RESID   2    AND NAME HA   ) (RESID   2    AND NAME HB   )   2.7  0.9  0.3
ASSIGN (RESID   2    AND NAME HA   ) (RESID   2    AND NAME HG2# )   2.2  0.4  1.8
ASSIGN (RESID   2    AND NAME HA   ) (RESID   2    AND NAME HG1# )   2.2  0.4  1.8
ASSIGN (RESID   2    AND NAME HA   ) (RESID   3    AND NAME HN   )   2.2  0.4  0.3
ASSIGN (RESID   3    AND NAME HB2  ) (RESID   3    AND NAME HD#  )   4.2  2.4  2.7
```

The first line means that we have a cross peak between the atom called HA in residue 1 and the atom called HN in residue number two. The interpretation of the following three numbers can be slightly different. Here, the upper bound is calculated as the sum of the first and the third number, and the lower bound is calculated as the difference between the first and the second number. If no stereospecific assignment could be made, or if the listed NOE is the average of several hydrogen atoms, this is indicated by a * or a # sign. Historically, the # was used for methylene and the * for methyl groups, but this convention is not always followed. The names of the atoms are usually according to the IUPAC conventions[9] with one obvious exception. Stereospecifically assigned β-hydrogen atoms, are according to the IUPAC conventions called $H^{\beta 2}$ and $H^{\beta 3}$, in the XPLOR format they are listed as HB2 and HB1, respectively! One should be absolutely sure whether any corrections have already been added to these distances or not. In the current example, the file already contains pseudo atom corrections according to ref [5], with the exception of types 6 and 7, where a correction of 2.4 Å was applied in stead of 2.9 Å and 2.0 Å, respectively.

In order to be able to read in the XPLOR format, we will need to modify the file slightly: just give seven columns with residue number and atom name of hydrogen atom I, residue number and atom name of hydrogen atom S and the three numbers that determine the NOE distance bounds. All * or # need to be replaced by the @ sign, because in GROMOS the # represents a comment and will be ignored upon reading. Also

add a title and put the NOE's in an NOESPEC block. The seven NOE's above should now look like:

```
TITLE
HEN LYSOZYME: 1e8l.mr
END
NOESPEC
    1     HA       2     HN      2.2  0.4  0.3
    2     HN       2     HA      4.2  2.4  0.3
    2     HA       2     HB      2.7  0.9  0.3
    2     HA       2     HG2@    2.2  0.4  1.8
    2     HA       2     HG1@    2.2  0.4  1.8
    2     HA       3     HN      2.2  0.4  0.3
    3     HB1      3     HD@     4.2  2.4  2.7
```

If one has the NOE intensities only in terms of strong, medium, weak or very weak, it should be relatively easy to bring your data in this format as well.

In order to interpret this file, the program prepnoe can be used. It reads in the NOE distances and together with a carefully created library file, determines for every proton involved in the NOE, the corresponding GROMOS type, as discussed in the previous section. In the case of virtual or pseudo atoms, it also determines which heavy atoms are needed to define their positions. For stereospecifically assigned methylene hydrogen atoms (type 4), the library file is able to assign them correctly for all standard amino acid side chains (HB1 = $H^{\beta 3}$ and HB2 = $H^{\beta 2}$). If this information is not added to the library, the program generates two distances to monitor, one for every atom. This means that the total number of NOE distances seems to increase, which needs to be corrected at a later point in the analysis.

The prepnoe program can also add corrections, which have been defined in a separate file. For every type we can apply a multiplicity correction and a pseudo atom correction. The program also offers the possibility to remove the corrections from the upper bounds, which will allow you to later apply another set of corrections.

Prepnoe produces two separate kinds of output. To stdout it write the DISRESSPEC block, which can be used in to actually calculate the NOE distances from a trajectory, or to apply distance restraints during the simulation. Its format is defined in the GROMOS manual, for the above example it looks like this. The corrected upper bound is written in the 11th column.

```
DISRESSPEC
# DISH: carbon-hydrogen distance
# DISC: carbon-carbon distance
# DISH,DISC
0.100 0.153
```

```
    5    3    6   14  1   17    0    0    0  0   0.250 1.0000
# 1 1LYSH CA HA  2VAL H HN
   17    0    0    0  0   18   16   19   22  1   0.450 1.0000
# 2 2VAL H HN  2VAL CA HA
   18   16   19   22  1   19   18   20   21  1   0.300 1.0000
# 3 2VAL CA HA  2VAL CB HB
   18   16   19   22  1   21   19    0    0  5   0.516 1.0000
# 4 2VAL CA HA  2VAL CG2 HG2@
   18   16   19   22  1   20   19    0    0  5   0.516 1.0000
# 5 2VAL CA HA  2VAL CG1 HG1@
   18   16   19   22  1   25    0    0    0  0   0.250 1.0000
# 6 2VAL CA HA  3PHE H HN
   27   26   28    0  4   28    0    0    0  7   1.010 1.0000
   27   28   26    0  4   28    0    0    0  7   1.010 1.0000
#automatically generated NOE distance to monitor!
# 7 3PHE CB HB2 3PHE CG HD@
```

Additionally prepnoe can write a filter file, which will allow you to post-process the calculated distances and remove the automatically added distances of type 4. The filter file contains an NOEFILTER block:

```
NOEFILTER
# mol residue atom atom mol residue atom atom     r0  filter    noe
1   1  1 LYSH   CA   HA   1  2 VAL     H   HN   0.25    1
2   1  2 VAL     H   HN   1  2 VAL    CA   HA   0.45    1
3   1  2 VAL    CA   HA   1  2 VAL    CB   HB    0.3    1
4   1  2 VAL    CA   HA   1  2 VAL   CG2 HG2@    0.4    1
5   1  2 VAL    CA   HA   1  2 VAL   CG1 HG1@    0.4    1
6   1  2 VAL    CA   HA   1  3 PHE     H   HN   0.25    1
7   1  3 PHE    CB  HB2   1  3 PHE    CG  HD@   0.69    2     8
8   1  3 PHE    CB  HB2   1  3 PHE    CG  HD@   0.69    2     7
```

In the last column, labeled filter, those NOE's that are were generated from a single experimental peak are linked together, so that at post processing time they can be reduced to a single value. The filter file allows for more post-processing options.

After preparing the input for the distance calculation, things get easier. For an analysis one can read in the DISRESSPEC block and calculate for the distances between the (virtual or pseudo) atoms. The time averaging of these distances is done both with p=3 and with p=6, so that one can later investigate the differences of both. The violations with respect to the upper bounds are also calculated and printed out, but one should be warned that this is possibly over the expanded set of data, with additional NOE distances.

This one can then remedy by using the program postnoe. It reads the filter file, and the output of the noe program. For those NOE distances that have been added to the set for type 4, it can collaps those back to a single value, to obtain the minimum of maximum violation. Note that this means that the HB2 in our example above, is not

automatically assigned to the $H^{\beta 2}$ in Phe, but rather to $H^{\beta 2}$ or $H^{\beta 3}$, which ever fits the experiment better (or worse). In order to do the assignment correctly, you will have to manually edit the filter file, replace the 2 in the 13th column by a one for the NOE distance that you want to keep, and with a 0 for the one you want to discard. By specifying 0 in this column, you can also remove other NOE's from your data set.

Furthermore the postnoe program allows you to decide whether to calculate the violations using a p=3 or p=6 averaging scheme, to remove NOE's that are longer than a specified distance, or to recalculate the violations from a different set of upper bound (*e.g.* obtained by applying different corrections).

For the output the postnoe program will print out the violations for all the NOE's as well as the average violation. This is defined as the sum of all violations, divided by the total number of NOE's in the set. Additionally it can print out a distribution of the violations.

Conclusion

Analyzing the NOE distance restraints is a mess! We have generously tried to provide some insight into the matter and some tools that make the work a bit easier. For this, we think we deserve eternal thanks.

References

1. Smith, L. J.; Sutcliffe, M. J.; Redfield, C. and Dobson, C. M. J Mol Biol 1993, 229, 930 - 944.
2. Schwalbe, H.; Grimshaw, S. B.; Spencer, A.; Buck, M.; Boyd, J.; Dobson, C. M.; Redfield, C. and Smith, L. Protein Science 2001, 10, 677 - 688.
3. Tropp, J. J Chem Phys 1980, 72, 6035 - 6043.
4. Fletcher, C. M.; Jones, D. N. M.; Diamond, R. and Neuhaus, D. Journal of Biomolecular NMR 1996, 8, 292 - 310.
5. Wüthrich, K.; Billeter, M. and Braun, W. J Mol Biol 1983, 169, 949 - 961.
6. Wüthrich, K. NMR of Protein and Nucleic Acids; John Wiley: New York, 1986.
7. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P. and Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.
8. Koning, T. M. G.; Boelens, R. and Kaptein, R. Journal of Magnetic Resonance 1990, 90, 111 - 123.

9. Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E. and Wüthrich, K. Journal of Biomolecular NMR 1998, 12, 1-23.