Supporting Information:

Assignment of Ile, Leu and Val Methyl Correlations in Supra-Molecular Systems: An Application to Aspartate Transcarbamoylase

Algirdas Velyvis[1], Howard K. Schachman[2] and Lewis E. Kay[1]*

[1]Departments of Biochemistry, Molecular Genetics, and Chemistry, University of Toronto, Toronto, ON, Canada M5S 1A8 and [2]Department of Molecular and Cell Biology and Virus Laboratory University of California, Berkeley, California 94720

Generation of suitable mutants for assignment via EXSY:

X-ray studies of ATCase establish that contacts between the r- and c-chains in the T state occur at what are termed r1-c1 and r4-c1 interfaces[1]. Since the r4-c1 interface plays a role in T state stabilization and in addition is less extensive than r1-c1, we focused exclusively on r1-c1 interactions. Residues Glu109, Glu117, Asp129, Ser131 and Asn132 in the c-chain and Asn111, Asn113, Glu119, Lys139, Glu142 and Lys143 in the r-chain were identified as candidates for removal, since they participate in inter-chain hydrogen bonds and salt bridges. Our hope was that removal of these interactions would increase the off-rate of r_2 from the complex, possibly to within the window where exchange spectroscopy would be of use. The first set of c-chain mutants included single substitutions, E109A and E117A, triple substitutions D129A/S131A/N132A and the quadruple mutant E117A/D129A/S131A/N132A, while r-chain mutants included N111A/N113A and K139A/E142A/K143A. The double r-chain mutant has significantly lower stability

compared to WT r_2, manifested by much lower expression yields and rapid precipitation of purified protein.

The impact of these mutations on holoenzyme dissociation is readily established by gel filtration chromatography, since the large difference in sizes between the 300 kDa enzyme, the 100 kDa c-chain trimer and the 30 kDa r-chain dimer lead to their baseline separation on a Superdex 200 column (black traces in Figure 1S). In an experiment where WT c-chain trimers are mixed with excess WT r-chain dimers two peaks are observed, corresponding to r_2 and r_6c_6. By contrast, if mutations in r- and c-chains are such that they no longer bind, the mixing/gel filtration experiment would result in a peak corresponding to excess r-chain dimers as before, with an additional peak corresponding to c_3 and no peak for r_6c_6. In cases where peak elution volumes are intermediate between what is observed for WT-r_6c_6 and c_3 they can be very qualitatively interpreted in terms of an enzyme dissociation rate: an elution profile close to what is observed for the c-chain trimer indicates a complex with a 'fast' off-rate, whereas a peak which elutes close to the WT-enzyme position is consistent with an off-rate sufficiently slow to allow complexes to persist for minutes at least. In order to quantify the elution profile we define an elution factor f as $f=(V_{mix}-V_{free})/(V_{WT,holo}-V_{WT,free})$, where V_{mix} is the elution volume of a mixture of r- and c-chains, V_{free} is the elution volume of free c-chains in absence of r-chains, $V_{WT,holo}$ and $V_{WT,free}$ are elution volumes of WT enzyme and WT c_3, respectively. Elution factors should range from 1 (same elution as WT-r_6c_6) to 0 (effectively no binding between r- and c-chains). Values of f outside the interval [0 1] reflect either experimental uncertainties in measurement of elution volumes or result from an 'abnormal' migration due to the effect of mutations. Repeated injections of the same r- and c-chain mixtures

lead to elution peaks within 0.1 mL, which, given that $V_{WT,holo}$-$V_{WT,free}$≈2 mL, predicts an uncertainty of about 5% in f. Table 1S lists f values for various permutations of mixtures of r- and c-chains that are generated with the initial set of mutations listed above. Notably, a mixture of the N111A/N113A r-chain and the E117A/D129A/S131A/N132A c-chain produced a 'complex' that elutes close to the free c-chain trimer, however, an f value of 0.21 indicates that binding is still not completely disrupted. Furthermore the instability of the double mutant r-chain is such that this construct is not an option for NMR studies.

A second iteration of mutagenesis was therefore carried out, using a quintuple c-chain mutant E109A/E117A/D129A/S131A/N132A ("5Ala c-chain") and single or double r-chain mutants, N111A, N113A, E119A, E142A or E119A/E142A. As shown in Table 1S and Figure 1S, the combination of the E119A/E142A r-chain with the 5Ala c-chain does not produce a stable complex (f ≈ 0). Moreover, by progressively adding back interacting side-chains a range of dissociation rates can be obtained, from 'fast' (E119A/E142A r-chain with 5Ala c-chain), to 'slower' (E119A r-chain, E117A/D129A/S131A/N132A c-chain) to very 'slow' (E119A r-chain, D129A/S131A/N132A c-chain), with an elution profile close to the WT complex, Figure 1S.

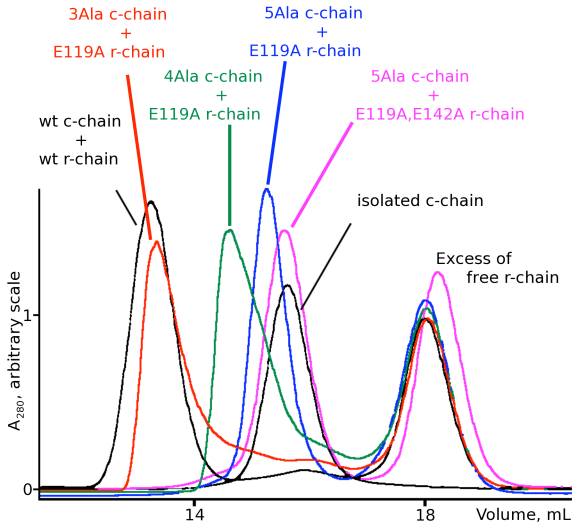

Figure 1S. Dissociation profiles of selected mutants of ATCase by gel filtration. Black traces indicate positions of WT ATCase, WT r-chain dimers and WT c-chain trimers. Elution traces from left to right correspond to the following ATCase mutants: red, E119A r-chain, D129A/S131A/N132A c-chain (f=0.95); green, E119A r-chain, E117A/D129A/S131A/N132A c-chain (f=0.42); blue, E119A r-chain, E109A/E117A/D129A/S131A/N132A c-chain (f=0.12); magenta, E119A/E142A r-chain, E109A/E117A/D129A/S131A/N132A c-chain (f=-0.02).

Table 1S. Characterization of binding between mutant r- and c-chains of ATCase by gel filtration chromatography.

Mutant set 1			Mutant sets 1 and 2		
c-chain	r-chain	f[a]	c-chain	r-chain	f[a]
WT	WT	1[b]	WT	E119A/E142A	0.88
WT	N111A/N113A	0.98	E117A	E119A	1.00
WT	K139A/E142A/K143A	1.00	E117A	E142A	0.98
E109A	K139A/E142A/K143A	0.95	E117A	E119A/E142A	0.13
E117A	WT	0.93	D129A/S131A/N132A	E119A	0.95
E117A	N111A/N113A	0.93	D129A/S131A/N132A	E142A	0.93
D129A/S131A/N132A	WT	0.94	D129A/S131A/N132A	E119A/E142A	0.08
D129A/S131A/N132A	N111A/N113A	0.84	E117A/D129A/S131A/N132A	E119A	0.42
D129A/S131A/N132A	K139A/E142A/K143A	0.81	E117A/D129A/S131A/N132A	E142A	0.32
E117A/D129A/S131A/N132A	WT	0.93	E117A/D129A/S131A/N132A	E119A/E142A	-0.02
E117A/D129A/S131A/N132A	N111A/N113A	0.21	E109A/E117A/D129A/S131A/N132A	WT	0.91
E117A/D129A/S131A/N132A	K139A/E142A/K143A	0.80	E109A/E117A/D129A/S131A/N132A	N111A	0.91
			E109A/E117A/D129A/S131A/N132A	N113A	0.82
			E109A/E117A/D129A/S131A/N132A	E119A	0.12
			E109A/E117A/D129A/S131A/N132A	E142A	0.15
			E109A/E117A/D129A/S131A/N132A	E119A/E142A	-0.02

[a] Defined above.

[b] By definition.

The PCS is the sum of contributions from two symmetric r-chains:

In the case of Co_2-ATCase the net pseudo-contact shift at a methyl position in one of the r_2 chains (denoted here by A or B), $^{PC}_{A}$, can be computed by adding contributions from A- and B-chain bound Co^{2+} ions, $^{PC}_{A} = {}^{PC}_{A,CoA} + {}^{P}_{A,CoB}$. In an asymmetrical molecule this would require the determination of two susceptibility tensors. However, since the two r-chains are related by a non-crystallographic dyad axis in the r-chain dimer[1], with identical metal coordination in both, the susceptibility tensors of the two metal ions must at least be very similar and related to each other by rotation about the dyad axis. Evidence supporting the fact that the two tensors are in fact identical comes from the observation of only a single set of resonances for methyl groups of each of the r-chains in the dimer

environment (both for r_2 and r_6c_6), suggesting that the structural properties of each chain are equivalent, and that only a single peak is observed for each methyl in Co_2-ATCase. Thus, the contribution of cobalt B to the position of a given methyl in chain A is identical to the contribution of cobalt A to the shift of the methyl in chain B, $^{PC}_{A,CoB} = {}^{PC}_{B,CoA}$, so that $^{PC}_{A} = {}^{PC}_{A,CoA} + {}^{PC}_{B,CoA}$ (see Materials and Methods). Given that the coordinates of atoms in chains A and B in the X-ray structure[2] are not identical (rmsd of 0.4 Å for C atoms in regions of regular secondary structure) we have calculated an average pseudo-contact shift as $^{PC} = (\,{}^{PC}_{A} + {}^{PC}_{B})/2 = (\,{}^{PC}_{A,CoA} + {}^{PC}_{A,CoB} + {}^{PC}_{B,CoA} + {}^{PC}_{B,CoB})/2$ and compared these values with experimental measurements to obtain methyl group assignments (see text).

A summary of the strengths and weaknesses of the methods described in the text:

We have described three complementary approaches for the assignment of methyl groups from very high molecular weight protein complexes. Although high quality methyl-TROSY spectra can be generated for many such systems the assignment of such spectra, necessary for further studies, remains non-trivial and must be approached on a case-by-case basis. It is unlikely that a general approach that performs equally well in all cases will emerge. Rather a 'multi-pronged' strategy is likely to be best with the details of optimization dependent on the system. The three methods considered in the text include: (1) 'Divide and conquer', which relies on the assignment of smaller fragments of the complex with the assignments subsequently transferred to the larger particle by comparing both through-bond and through-space spectra. This approach has been used with success in studies of the proteasome[3] and SecA[4]. (2) Exchange spectroscopy, in

which assignments corresponding to the 'free' state of a fragment/subunit of a larger complex are transferred to the 'bound' state by magnetization exchange. Applications of this approach have to this point being restricted to relatively small protein complexes, with exchange parameters that naturally fall into the slow exchange window of the EXSY technique[5,6]. (3) Measurement of PCSs. Such values have been used as structural tools for many decades in protein NMR[7] and more recently they have found application in assignment of methyl groups in relatively small proteins for which the susceptibility tensor is already known from ^{15}N-^{1}H correlation spectroscopy[8].

It is of interest to compare the three assignment methods highlighted in this study. It is clear that implementation of all three approaches will very likely require some level of cloning and protein engineering beyond that needed for the production of a methyl labeled sample for recording an initial methyl-TROSY spectrum. This is certainly the case for the 'divide and conquer' approach in which low molecular weight fragments are produced, as exemplified by studies of SecA[4] and the proteasome[3], although the situation with ATCase is somewhat simplified by virtue of the fact that the isolated r-chain dimer is stable in solution. Assignments via EXSY require production of a small fragment that then exchanges with the larger complex in the slow exchange regime. This, in general, will require substantial mutagenesis to tune the thermodynamics and the kinetics of complex formation appropriately. Studies involving PCS measurements potentially require the least amount of protein engineering, ranging from none if there is a natural metal site in the protein, to the creation of a few point mutants for the introduction of cysteines (into a cysteine free background) for site-specific tagging. It is, of course, desirable to minimize the amount of mutagenesis that must be performed. Mutations can

lead to significant chemical shift changes that complicate the transfer of chemical shifts from smaller to larger particles ('divide-and-conquer') or between mutant and WT complexes (EXSY or PCSs).

The methods also have differing requirements for *a priori* structural information. The NOE analysis that forms the basis for the 'divide-and-conquer' approach can, in principle, be performed in the absence of a structure if the NOE patterns in the assigned fragment and in the large particle are very similar. However, in our experience a high quality X-ray structure of the complex is essential for the correct interpretation of at least some of the NOEs used for assignment. In the present study a 2.1 Å structure of the T-state of ATCase[2] (PDB code 1TUG) was used. By contrast, assignment via EXSY does not rely on any structural information whatsoever for data interpretation, although some knowledge of structure greatly facilitates the choice of mutations that are necessary for creating an exchanging system in the first place. Finally, the PCS approach is heavily tied to an accurate structure. This was made clear in the present study by a comparison of RMSD values quantifying differences between measured PCSs and those predicted based on two nearly equivalent ATCase T-state structures refined to 2.1 Å (pdb code 1TUG)[2] and 2.2 Å (pdb code 1ZA1)[9]; RMSD values were as much as a factor of 4 different, with the 1ZA1 structure always giving poorer results. This implies that if possible several structures should be tested in analyses involving PCSs.

An additional complication in the interpretation of PCS values arises in the case where the location of the paramagnetic center is not known precisely. For systems like ATCase containing a high affinity metal binding site whose position is well defined by crystallography this is not an issue. However, for larger coordinating ligands such as

EDTA[10] or peptide tags[11] attached to the protein surface by disulfides the situation is less favorable. There has been recent progress in this regard, however, including the development of a dipicolinic acid tag for binding lanthanides, recently proposed by Otting and coworkers, that is smaller and more rigid[12]. Another exciting possibility is to site-specifically incorporate small, rigid non-natural metal chelating amino acids into the protein of interest via stop-codon suppression[13,14]. In cases where dynamics of the metal bound tag are an issue they can be partially taken into account by fitting PCS values to 8 parameters that include the 5 values required for defining the susceptibility tensor, with 3 additional values defining the average position of the metal. Although not considered in the present study, the measurement of PCS values and perhaps paramagnetic relaxation enhancements from paramagnets attached to several positions (one at a time) provides a powerful approach for breaking degeneracies in assignment that arise when two peaks have identical PCS values from a single tag.

Extensive use of 'assignment via mutation' has not been made for ATCase here. However, studies of other complexes that are in progress in our laboratory have exploited this simple approach. As mentioned above, and described previously in connection with our work on the proteasome[3], care must be taken in the interpretation of spectra that derive from single point mutants because often there are changes in the positions of many peaks, especially when a number of methyl containing residues are found in the 'mutated' cluster. However, by comparing spectra of mutants that derive from the same structural region it may be possible in some cases to resolve ambiguities that derive from secondary influences of the mutation, as was done here. Finally, computational approaches for predicting chemical shifts from 3D structural models are improving and considerable

success has been reported in the assignment of the Ile/Leu/Val methyl groups of the - rings of the proteasome based on an automated approach that incorporates X-ray structural information and 3D NOESY/through-bond data sets[15].

Table 2S. Summary of approximate times for sample production and experiment acquisition

Method	Number of DNA constructs, mutants prepared	Time in wet lab to optimize conditions, prepare samples [(a)]	NMR samples [(b)]	NMR experiments	NMR spectrometer time
Divide and conquer	0 [(c)]	6 weeks	a) U-^{2}H,^{15}N,^{13}C, ILV r-dimer b) U-^{2}H,^{15}N,^{13}C, ILV holoenzyme c) U-^{2}H, ILV r-dimer d) U-^{2}H, ILV holoenzyme	<u>r-dimer:</u> HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HNCO, HN(CA)CO, HMCM[CG]CBCA, Ile, Leu-HMCM(CGCBCA)CO, Val-HMCM(CBCA)CO, ^{13}C/^{13}C-separated NOESY <u>holoenzyme:</u> HMCMCG/CB (at 37 and 45 °C), HMCM(CG/CB)CB/CA, Ile,Leu-HMCM(CGCB)CA, ^{13}C/^{13}C-separated NOESY (at 37 and 45 °C)	15 days 14 days
EXSY [(d)]	13	12-14 weeks	a) U-^{2}H, Ile 1 rE119A r-dimer b) U-^{2}H, LV rE119A r-dimer Both samples were later converted into a partially or fully c-chain bound state.	EXSY spectra, HMQC	2.75 days
PCS	0 [(c)]	6-8 weeks	a) U-^{2}H, Ile 1 holoenzyme b) U-^{2}H, ILV holoenzyme	HMQC @ 800 MHz and 500 MHz	1.5 days [(e)]
Muta-genesis	5	1 week [(f)]	U-^{2}H, Ile 1 holoenzyme	1 HMQC / mutant	30 min / mutant

(a) For all methods (excepting mutagenesis) the dominant fraction of wet laboratory time (1/2 - 2/3) was spent in optimizing techniques and conditions. Once these one-time

'costs' are absorbed, experiments on samples with different labeling pattern, or on closely related molecular systems can be repeated much faster.

(b) Labeling of methyl groups in NMR samples: "Ile 1" labeling: all 1 methyl groups of isoleucines are of the $^{13}CH_3$ variety with no labeling of leucines and valines;
"ILV" labeling: all 1 methyl groups of isoleucines are of the $^{13}CH_3$ variety, leucine and valine methyl groups are labeled with $^{13}CH_3$ at one of methyl groups while the other is $^{12}CD_3$, (distribution of labeled and unlabeled methyls at the prochiral centre is random);
"LV" labeling: as ILV, but no labeling of isoleucines.

(c) The situation is fortuitous for ATCase, as no molecular biology is required to divide the holoenzyme into smaller components or to introduce paramagnetic metal into the molecule. Substantial investments in time may be required for other systems.

(d) List of NMR samples, experiments and spectrometer time for assignment by exchange assumes that assignments of isolated r-dimer are already available.

(e) Long acquisition time HMQC spectra are needed to observe as many peaks as possible due to paramagnetic broadening.

(f) If the number of mutants is not too large they can be prepared in parallel as was done here.

References:

(1) Honzatko, R. B.; Crawford, J. L.; Monaco, H. L.; Ladner, J. E.; Ewards, B. F.; Evans, D. R.; Warren, S. G.; Wiley, D. C.; Ladner, R. C.; Lipscomb, W. N. *J. Mol. Biol.* **1982**, *160*, 219-263.

(2) Stieglitz, K.; Stec, B.; Baker, D. P.; Kantrowitz, E. R. *J. Mol. Biol.* **2004**, *341*, 853-868.

(3) Sprangers, R.; Kay, L. E. *Nature* **2007**, *445*, 618-622.

(4) Gelis, I.; Bonvin, A. M.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G. *Cell* **2007**, *131*, 756-769.

(5) Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M., Skelton, N.J. *Protein NMR Spectroscopy: Principles and Practice*; 2nd ed.; Academic Press, 2007.

(6) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546-4553.

(7) Bertini, I.; Luchinat, C.; Piccioli, M. *Methods Enzymol.* **2001**, *339*, 314-340.

(8) John, M.; Schmitz, C.; Park, A. Y.; Dixon, N. E.; Huber, T.; Otting, G. *J. Am. Chem. Soc.* **2007**, *129*, 13749-13757.

(9) Wang, J.; Stieglitz, K. A.; Cardia, J. P.; Kantrowitz, E. R. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8881-8886.

(10) Gaponenko, V.; Sarma, S. P.; Altieri, A. S.; Horita, D. A.; Li, J.; Byrd, R. A. *J. Biomol. NMR* **2004**, *28*, 205-212.

(11) Ma, C.; Opella, S. J. *J. Magn. Reson.* **2000**, *146*, 381-384.

(12) Su, X. C.; Man, B.; Beeren, S.; Liang, H.; Simonsen, S.; Schmitz, C.; Huber, T.; Messerle, B. A.; Otting, G. *J. Am. Chem. Soc.* **2008**, *130*, 10486-10487.

(13) Xie, J. M.; Liu, W. S.; Schultz, P. G. *Angew. Chem. Int. Edit.* **2007**, *46*, 9239-9242.

(14) Lee, H. S.; Spraggon, G.; Schultz, P. G.; Wang, F. *J. Am. Chem. Soc.* **2009**, *131*, 2481-2483.

(15) Xu, Y.; Liu, M.; Simpson, P. J.; Isaacson, R.; Cota, E.; Marchant, J.; Yang, D.; Zhang, X.; Freemont, P.; Matthews, S. *J. Am. Chem. Soc.* **2009**, *131*, 9480-9481.