

values has indicated that for BPR and CatV ligands, chelation of $[\text{Co}(\text{1,10-phen})_2]^{2+}$ moiety occurs from 3', 4' position of the quinone ring (see Scheme 4). Hence, on the TiO_2 surface anchoring process takes place through the gallol and catechol rings of the BPR and CatV ligands respectively. The surface anchoring observed for the two complexes is quite similar to the anchoring which occurs in the case of anthrocyanine dye material².

The findings clearly suggest that non-planarity of the triphenylmethane-type ligand strongly influences the observed photocurrent conversion efficiencies. This is reflected by the different intensities of the observed MLCT transitions for the complexes. Variation in the strength of the MLCT transition occurs as a result of different torsion angles. In the case of high torsion angles substantial $d\pi-\pi^*$ interactions take place, and similarly, in the case of low torsion angles less $d\pi-\pi^*$ interactions occur.

1. Desilvestro, J., Gratzel, M., Kavan, L. and Moser, J., *J. Am. Chem. Soc.*, 1985, **107**, 2988–2992.
2. (a) Tennakone, K., Kumara, G. R. R. A., Kottegoda, I. R. M., Wijayantha, K. G. U. and Perera, V. P. S., *J. Phys. D: Appl. Phys.*, 1998, **31**, 1492–1495; (b) Nazeeruddin, M. K. *et al.*, *J. Am. Chem. Soc.*, 1993, **115**, 6382–6386.
3. Cherepy, N. J., Smestad, G. P., Gratzel, M. and Zhang, J. Z., *J. Phys. Chem. B*, 1997, **101**, 9342–9345.
4. Jayaweera, P. M., Kumarasinghe, A. R. and Tennakone, K., *J. Photochem. Photobiol. A*, 1999, **126**, 111–114.
5. Jayaweera, P. M., Palayangoda, S. S. and Tennakone, K., *J. Photochem. Photobiol. A*, 2001, **140**, 173–176.
6. Burns, D. T. and Dadger, D., *Analyst*, 1980, **105**, 1082–1086.
7. Vinodgopal, K., Hua, X., Dahlgren, R. L., Lappin, A. G., Patterson, L. K. and Kamat, P. V., *J. Phys. Chem.*, 1995, **99**, 10883–10887.
8. Zemer, M. C., *Inorg. Chem.*, 1986, **28**, 2728–2733.

ACKNOWLEDGEMENT. We are grateful to the National Science Foundation for support.

Received 20 March 2002; revised accepted 25 September 2002

Automated NMR assignments of proteins for high throughput structure determination: TATAPRO II

H. S. Atreya, K. V. R. Chary* and Girjesh Govil

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India

TATAPRO (Tracked Automated Assignments in Proteins), a novel algorithm for automated NMR assignments in proteins is presented to aid in high throughput protein structure determination. In this version (TATAPRO II), the 20 amino acid residues are classified into nine distinct categories based on their characteristic $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts derived statistically using a database of ~100,000 shifts. Further, in the current version, the N- and C-terminal residues of an assigned polypeptide stretch are retained during the course of resonance assignments. This results in faster execution time of the program, increased efficiency and greater robustness towards missing peaks.

SEQUENCE-specific resonance assignment (hereafter abbreviated as *ssr_assignments*) of all the NMR-active nuclei in a protein constitutes the first step towards the complete characterization of its three-dimensional structure¹. *Ssr_assignments*, if carried out manually, constitute a tedious and time-consuming task. For this reason, there

have been several efforts to automatize the assignment process². With the advent of multidimensional, triple-resonance (^1H , ^{13}C , ^{15}N) strategies for resonance assignments, it has become increasingly clear that the information content of protein spectra can allow complete automation of *ssr_assignments*³. This advance has tremendous implications for the growth of NMR spectroscopy as a powerful tool in structural genomics^{4,5}, which involves high-throughput protein structure determination.

A number of strategies have been proposed for *ssr_assignments* in proteins². These include approaches which utilize information from triple-resonance experiments and methods such as simulated annealing⁶, Bayesian statistics and artificial intelligence⁷, characteristic $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts of individual residues⁸, threshold-accepting algorithm⁹, connectivity-tracing algorithms¹⁰ and neural networks¹¹.

We have recently proposed a novel algorithm for such automated assignments called TATAPRO (Tracked Automated Assignments in Proteins)¹². TATAPRO achieves *ssr_assignments* of $^1\text{H}^\text{N}$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'/^1\text{H}^\alpha$ and ^{15}N spins by utilizing the protein primary sequence and peak lists from a set of 3D triple-resonance spectra¹², namely CBCANH¹³, CBCA(CO)NH¹⁴, HNCO¹⁵ and HN(CA)CO¹⁶. Peak lists obtained from these spectra consisting of chemical shift coordinates of the peaks, $(\omega_1, \omega_2, \omega_3) = (^{13}\text{C}/^1\text{H}^\alpha, ^{15}\text{N}, ^1\text{H}^\text{N})$, along with their intensities and phases are taken as inputs for TATAPRO. For example, automatically picked CBCA(CO)NH peak list has information about $^{13}\text{C}_{i-1}^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shifts for a given pair of $^{15}\text{N}_i$ and $^1\text{H}_i^\text{N}$. From such a list, the chemical shifts

*For correspondence. (e-mail: chary@tifr.res.in)

of $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_{i-1}^\beta$ are identified for each specific pair of $^{15}\text{N}_i$ and $^1\text{H}_i^\text{N}$ chemical shifts within the user-defined tolerance limits and grouped into a single set. Likewise, automatically picked CBCANH peak list has information about $^{13}\text{C}_i^\alpha$, $^{13}\text{C}_i^\beta$, $^{13}\text{C}_{i-1}^\alpha$ and $^{13}\text{C}_{i-1}^\beta$ chemical shifts, for a given pair of $^{15}\text{N}_i$ and $^1\text{H}_i^\text{N}$ chemical shifts. On the other hand, automatically picked HN(CA)CO and HNCO peak lists provide information about $^{13}\text{C}_i'$ and/or $^{13}\text{C}_{i-1}'$ chemical shifts. Besides, the success of TATAPRO relies on the classification of the 20 amino acid residues into eight different categories, which is based on their characteristic $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shifts, derived statistically from a large database of chemical shifts in the BioMagResBank (BMRB)¹⁷. The program has been successfully tested for resonance assignments using experimental data in three different proteins in the molecular weight range of 15–20 kDa (ref. 12). The program has also been tested for its robustness using published assignment data of four other proteins in the molecular weight range of 18–42 kDa (ref. 12).

During the course of its application to various different proteins, it was felt that the efficiency and speed of TATAPRO could be further improved by incorporating several changes in the program. The software has been modified accordingly, which has resulted in its increased robustness towards spectral overlap and missing data, a feature that is common when dealing with large molecular weight proteins. In this communication, we provide a detailed description of the improved version TATAPRO II and the results obtained thereby. The improved version of the program, called TATAPRO II, is available on request from chary@tifrr.res.in or through the BMRB web server: www.bmrb.wisc.edu.

An extensive statistical analysis was carried out using the $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shift information of all proteins currently deposited in the BMRB. Out of 1832 proteins for which chemical shifts are available, ^{13}C shifts are available for 839 proteins in the molecular weight range of 10–42 kDa. The total number of $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shifts reported so far is 58,768 and 45,463 respectively. The total number of amino acid residues for which both $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ shifts are available is 42,629. The distribution of chemical shifts for individual amino acid residues is given in Table 1.

Figure 1 shows a 2D plot of $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shifts. As is evident from the figure, one can classify amino acid residues into eight distinct categories (as against seven done earlier¹²) based entirely on the characteristic $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shifts (each of these categories is shown with a different colour): (i) Gly having no $^{13}\text{C}_i^\beta$; (ii) Ala having $14 < ^{13}\text{C}_i^\beta < 24$ ppm; (iii) Arg, Cys^{red}, Gln, Glu, His, Lys, Met, Val, Trp having $^{13}\text{C}_i^\beta$ in the range 24–36 ppm; (iv) Asp, Asn, Cys^{oxd}, Ile, Leu, Phe and Tyr having $^{13}\text{C}_i^\beta$ in the range 36–52 ppm; (v) Ser having $^{13}\text{C}_i^\beta$ in the range 56–67 ppm; (vi) Thr having $^{13}\text{C}_i^\beta > 67$ ppm; (vii) some Val residues having $^{13}\text{C}_i^\alpha > 64$ ppm and

$24 < ^{13}\text{C}_i^\beta < 36$ ppm, and (viii) some Ile residues having $^{13}\text{C}_i^\alpha > 64$ ppm and $36 < ^{13}\text{C}_i^\beta < 52$ ppm. Pro residues, which lack an amide proton ($^1\text{H}^\text{N}$), are not observed in any of the aforementioned triple resonance spectra and hence are classified as the ninth category. In our algorithm, each of these nine categories is distinguished from one another

Table 1. Distribution of $^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$ chemical shifts for individual amino acid residues used in the statistical analysis. Pro residues are excluded from the analysis as they do not show up in any of the triple-resonance spectra considered in the present algorithm (see text)

Amino acid	Number of chemical shifts		
	$^{13}\text{C}_i^\alpha$	$^{13}\text{C}_i^\beta$	$^{13}\text{C}_i^\alpha$ and $^{13}\text{C}_i^\beta$
Alanine	4617	4006	3790
Arginine	2935	2379	2102
Asparagine	2590	2181	1978
Aspartic acid	3832	3275	2978
Cysteine ^{oxd}	460	334	299
Cysteine ^{red}	572	506	490
Glutamic acid	2561	2120	2019
Glutamine	4834	4092	3891
Glycine	4396	0	0
Histidine	1305	1085	997
Isoleucine	3284	2787	2553
Leucine	5216	4336	4213
Lysine	4651	3890	3598
Methionine	1355	1090	1009
Phenylalanine	2319	1950	1887
Serine	3669	2976	2812
Threonine	3382	2761	2689
Tryptophan	686	580	456
Tyrosine	1939	1592	1456
Valine	4165	3523	3412
Total	58,768	45,463	42,629

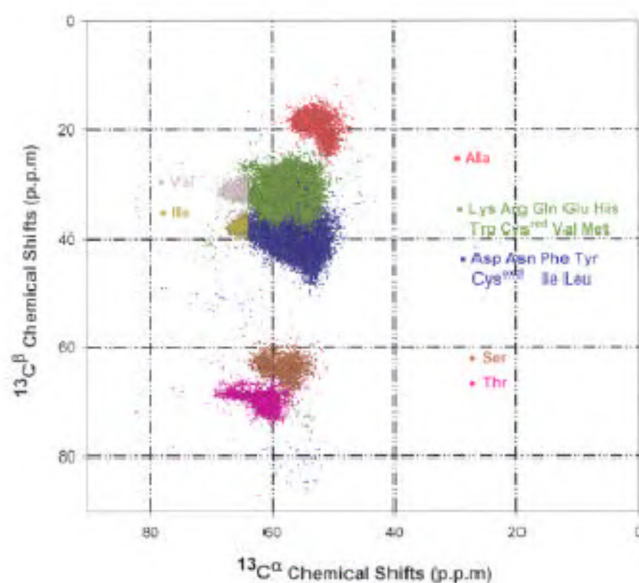


Figure 1. Two-dimensional plot of 42,629 $^{13}\text{C}_i^\alpha$ vs $^{13}\text{C}_i^\beta$ chemical shifts obtained from 839 proteins. The chemical shifts have been grouped into 7 categories, each shown with a different colour, based on the characteristic $^{13}\text{C}_i^\alpha$ vs $^{13}\text{C}_i^\beta$ shifts (see text).

by a two-digit code. In the earlier version of TATAPRO, amino acid residues were classified into eight different groups (Pro residues as the eighth category), with both Ser and Thr put under the same category¹². Their separation into separate categories in the current version of the program has resulted in accomplishing the assignments in relatively less number of iterations and with an increased percentage of assignments as discussed below.

Table 2 shows the different categories of amino acid residues described above. The last two columns in Table 2 indicate percentages of residues that violate the two-digit code assigned to them. This happens only if a given residue exhibits unusual $^{13}\text{C}^\alpha$ or/and $^{13}\text{C}^\beta$ chemical shift(s) as a result of which, it acquires a code different from the one generally expected. As is evident from Table 2, no dramatic differences have been observed in the percentage of $^{13}\text{C}^\beta$ chemical shift violations with the increased amount of chemical shift information, which is presently available in the BMRB. Interestingly, most of the violations are relatively less compared to what has been found earlier¹² (shown in parenthesis). The only exceptions are Ser and Thr, as they are presently given different two-digit codes. This is primarily due to the partial overlap of their $^{13}\text{C}^\beta$ chemical shifts. However, such an overlap does not have any adverse effect on our assignment procedure. On the contrary, as discussed below, the algorithm becomes more efficient and robust towards missing peaks.

Ssr_assignments using TATAPRO II can be divided into four essential steps: (i) preparing the chemical shift list (*master list*). This list contains individual sets or rows of $^1\text{H}_i$, $^{15}\text{N}_i$, $^{13}\text{C}_i^\alpha$, $^{13}\text{C}_i^\beta$, $^{13}\text{C}_i'$, $^{13}\text{C}_{i-1}^\alpha$, $^{13}\text{C}_{i-1}^\beta$, and $^{13}\text{C}_{i-1}'$ chemical shifts; (ii) assigning two-digit codes to the individual rows in the *master list* using the criteria shown in Table 2; (iii) starting assignments; this results in the creation of an *assign_array* and *pps_array* (described below); and (iv) mapping of the *assign_array* onto the *pps_array* for final ssr_assignments.

To begin with, the program reads in the $^{13}\text{C}_i^\alpha$, $^{13}\text{C}_i^\beta$, and $^{13}\text{C}_i'$ chemical shift values from the first row in the *master_list* and searches for a row where, within the user-defined tolerance limits, these three shifts are seen as $^{13}\text{C}_{i-1}^\alpha$, $^{13}\text{C}_{i-1}^\beta$ and $^{13}\text{C}_{i-1}'$ chemical shifts. If the search is successful and unique, the two-digit code associated with the new row is stored in an *assign_array*. This procedure corresponds to forward assignment in the primary sequence, which is continued till a break is encountered. The break can be due to a Pro residue, a missing peak(s) or if the C-terminal end of the polypeptide chain has been reached. Once a stretch of amino acid residues has been assigned in the forward direction, the program continues with the assignment in the backward direction, starting again from the first row in the *master_list*. For backward assignment, the program reads in the $^{13}\text{C}_{i-1}^\alpha$, $^{13}\text{C}_{i-1}^\beta$, and $^{13}\text{C}_{i-1}'$ chemical shifts for a given row in the *master_list*, and searches for the row where these chemical shifts are seen as $^{13}\text{C}_i^\alpha$, $^{13}\text{C}_i^\beta$ and $^{13}\text{C}_i'$ chemical shifts. If the search is successful and unique, the two-digit code associated with the new row is stored in the same *assign_array*, as was done in the case of forward assignment. The assignment is continued till a break is encountered. During this procedure, if more than one possible pair of $^1\text{H}_i$ and $^{15}\text{N}_i$ chemical shifts satisfies the assignment condition, the program continues with the assignment along each possible pathway until a break is encountered. After assigning a stretch of amino acid residues in both forward and backward directions, the program maps the *assign_array*(s) thus obtained onto the *pps_array* as the final step of ssr_assignments.

In the older version of TATAPRO, once a stretch of amino acid residues was assigned and before commencing with the next round of assignments, assigned rows and residues were deleted from the *master_list* and *pps_array* respectively. However, this resulted in the loss of crucial information, particularly for residues which constitute the N- and C-terminal ends of the assigned

Table 2. Two-digit code assigned to different amino acid residues based on their characteristic $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift ranges obtained using the most recent database of chemical shifts in the BMRB

$^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift (δ in ppm) characteristics	Amino acid	Two-digit code	Percentage of $^{13}\text{C}^\beta$ chemical shift violations*	Percentage of other residues taking the code*
Absence of $^{13}\text{C}^\beta$	Gly	1 0	0.0 (0.0)	0.0 (0.0)
$14 < \delta(^{13}\text{C}^\beta) < 24$	Ala	2 0	2.29 (0.8)	0.09 (0.09)
$56 < \delta(^{13}\text{C}^\beta) < 67$	Ser	3 0	3.06 (0.5)	0.38 (0.04)
$24 < \delta(^{13}\text{C}^\beta) < 36$ and $\delta(^{13}\text{C}^\alpha) < 64$	Lys, Arg, Gln, Glu, His, Trp, Cys ^{oxd} , Val and Met	4 0	3.16 (3.1)	2.59 (2.4)
$24 < \delta(^{13}\text{C}^\beta) < 36$ and $\delta(^{13}\text{C}^\alpha) \geq 64$	Val	4 1	5.02 (1.3)	2.57 (0.6)
$36 < \delta(^{13}\text{C}^\beta) < 52$ and $\delta(^{13}\text{C}^\alpha) < 64$	Asp, Asn, Phe, Tyr, Cys ^{oxd} , Ile and Leu	5 0	3.16 (3.0)	4.20 (2.4)
$36 < \delta(^{13}\text{C}^\beta) < 52$ and $\delta(^{13}\text{C}^\alpha) \geq 64$	Ile	5 1	8.04 (6.8)	1.85 (0.3)
—	Pro	6 0	—	—
$\delta(^{13}\text{C}^\beta) > 67$	Thr	7 0	5.25 (0.5)	0.37 (0.04)

*Numbers in parentheses indicate the percentage of violations obtained with the older version of TATAPRO.

Table 3. Details of test proteins and percentage of assignments obtained in each case using the old and the new versions of TATAPRO

Protein	Molecular weight (kDa)	Percentage of assignments obtained on random deletion of peaks*		
		0	15	30
<i>Drosophila</i> numb phosphotyrosine-binding domain	17.8	100	81.9 (80.1)	67.1 (54.5)
Fibroblast collagenase	18.7	100	71.3 (70.6)	61.4 (55.5)
<i>Borrelia burgdorferi</i> OspA	28	100	80.6 (80.2)	60.6 (58.5)
<i>Escherichia coli</i> maltose-binding protein	42	100	71.6 (70.7)	61.4 (57.2)

*Numbers in parentheses indicate the percentage of assignments obtained in test proteins using the older version of TATAPRO.

polypeptide stretch. These residues, if retained in the *master_list* and *pps_array*, could be used as markers. Hence, they help to assign those polypeptide stretches which lie in the immediate neighbourhood (in their N- or C-terminal sides) of the already assigned polypeptide stretches. Moreover, end-residues also serve as markers in resolving the ambiguity arising when the *assig_array* maps to more than one stretch in the *pps_array*. In light of this, in TATAPRO II, the N- and C-terminal residues of an assigned polypeptide stretch are retained in both the *master_list* and the *pps_array*. However, these amino acid residues are given a different two-digit code to distinguish them from those that are yet to be assigned. For example, an Ala that has been assigned and simultaneously lies either at the N- or the C-terminus of an assigned polypeptide stretch, is given a new code **2 1** in the *master_list* and *pps_array*. This distinguishes it from other unassigned Ala residues in the *master_list* and *pps_array* that have a code **2 0** (see Table 2). Such inclusion of the end-residues resulted in improved efficiency of the program. The program in the present form requires less number of iterations to find the correct *assig_array*, resulting in an increased speed of execution.

Table 3 summarizes the results obtained for *ssr_assignments* in the test proteins using TATAPRO II. With no random deletion of rows from the *master_list*, 100% assignments are obtained using both old and new versions of the program. This is because the initial data set is obtained using the published assignment of all the amino acid residues, and hence is perfect. However, the robustness of the program is tested on random deletion of rows from the *master_list*, which mimics a real situation wherein one encounters missing peaks. Even after random deletion of rows up to 30% from the *master_list*, TATAPRO II gives increased percentage of assignments in all the test proteins compared to the older version of the program.

Another modification concerns residues in an assigned polypeptide stretch that have unusual chemical shifts and therefore do not belong to their respective category. In such a situation (which we refer to as a 'mismatch'), the

mapping of *assig_array* onto the *pps_array* will result in either an incorrect mapping or no mapping. In the older version of the program, all residues were equally considered as candidates for a mismatch. However, refined statistics carried out using the latest database of chemical shifts reveals that Gly, Ala, Ser and Thr have the least $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift violations (Table 2). Hence, these residues rarely have mismatches. In light of this, in TATAPRO II, these residues are not considered to be mismatches during any stage of assignments. This substantially reduces the errors which occur if these residues are considered as mismatches. Further, during all stages of the assignments in TATAPRO II, a maximum of only one mismatch is allowed for mapping *assig_array* onto the *pps_array*. This is unlike the older version, wherein a maximum of three mismatches were allowed. This further decreased the chances of incorrect mapping. However, such criterion did not influence the final percentage of assignments obtained.

In conclusion, the modifications incorporated in TATAPRO II include refinement of the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift statistics and retaining the N- and C-terminal residues in an assigned polypeptide stretch in both the *master_list* and *pps_array*. This results in less number of iterations during the entire assignment procedure and hence, faster execution time of the program and concomitant increased efficiency in *ssr_assignments*. TATAPRO II has been tested using published data on four proteins in molecular weight range of 18–42 kDa, and shows increased percentage of assignments compared to the earlier version of TATAPRO.

1. Wüthrich, K., *NMR of Proteins and Nucleic Acids*, John Wiley, New York, 1986.
2. Moseley, H. N. B. and Montelione, G. T., *Curr. Opin. Struct. Biol.*, 1999, **9**, 635–642.
3. Ikura, M., Kay, L. E. and Bax, A., *Biochemistry*, 1990, **29**, 4659–4667.
4. Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C. and Szyperski, T., *Nature Struct. Biol. (Suppl.)*, 2000, **7**, 982–985.

5. Chary, K. V. R. and Atreya, H. S., *J. Postgrad. Med.*, 2002, **48**, 83–87.
6. Buchler, N. E. G., Zuiderwig, E. R. P., Wang, H. and Goldstein, R. A., *J. Magn. Reson.*, 1997, **125**, 34–42.
7. Zimmerman, D. E. *et al.*, *J. Mol. Biol.*, 1997, **269**, 592–610.
8. Friedrichs, M. S., Mueller, L. and Wittekind, M., *J. Biomol. NMR*, 1994, **4**, 703–726.
9. Leutner, M., Gschwind, R. M., Liermann, J., Schwarz, C., Gemmecker, G. and Kessler, H., *ibid*, 1998, **11**, 31–43.
10. Olson, J. B. Jr. and Markley, J. L., *ibid*, 1994, **4**, 385–410.
11. Hare, B. J. and Prestegard, H., *ibid*, 1994, **4**, 35–46.
12. Atreya, H. S., Sahu, S. C., Chary, K. V. R. and Govil, G., *ibid*, 2000, **17**, 125–136.
13. Wittekind, M. and Mueller, L., *J. Magn. Reson.*, 1993, **B101**, 201–205.
14. Grzesiek, S. and Bax, A., *J. Am. Chem. Soc.*, 1992, **114**, 6291–6293.
15. Kay, L. E., Ikura, M., Tschudin, R. and Bax, A., *J. Magn. Reson.*, 1990, **89**, 496–514.
16. Clubb, R. T., Thanabal, V. and Wagner, G., *ibid*, 1992a, **97**, 213–217.
17. Seavey, B. R., Farr, E. A., Westler, W. M. and Markley, J. L., *J. Biomol. NMR*, 1991, **1**, 217–236.

ACKNOWLEDGEMENTS. The facilities provided by the National Facility for High Field NMR, supported by Department of Science and Technology, Department of Biotechnology, Council of Scientific and Industrial Research, and Tata Institute of Fundamental Research, Mumbai are gratefully acknowledged. The help provided by two summer students, Mr Ajay Sheth and Mr Vikas Yadav, in the statistical analysis of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift information, is gratefully acknowledged.

Received 19 June 2002; revised accepted 16 October 2002

Mechanism of artificial transformation of *E. coli* with plasmid DNA – Clues from the influence of ethanol

Suchitra Sarkar, Sujan Chaudhuri and Tarakdas Basu*

Department of Biochemistry and Biophysics, University of Kalyani, Kalyani 741 235, India

The standard method of transformation of *E. coli* with plasmid DNA involves two important steps – binding of DNA to the cell surface, suspended in 100 mM CaCl_2 at 0°C , and the subsequent entry of DNA to the cell cytosol by a heat-pulse from 0 to 42°C . When competent *E. coli* cells were transformed with plasmid DNA in the presence of different concentrations (up to 10% v/v) of ethanol, the transformation efficiency (TR_E) decreased gradually with increase in ethanol concentration. This decrease in TR_E was directly proportional to ethanol-mediated leaching of lipopolysaccharide (LPS) molecules from the competent cell surface, indicating LPS was the major target site for DNA adsorption to the competent cells. *In vitro* spectrophotometric study showed evidence that there was binding interaction between plasmid DNA and *E. coli* LPS in the presence of a divalent cation, Ca^{2+} . Moreover, plasmid DNA, previously incubated with LPS in CaCl_2 , had less ability to transform *E. coli* cells. The results suggest that during artificial transformation of *E. coli*, the naked DNA was first bound to the LPS molecules on the competent cell surface and uptake of this LPS-absorbed DNA into the cell cytosol was associated with CaCl_2 -mediated cell-membrane disintegration.

THE technique of DNA transformation has become important in virtually all aspects of molecular genetics. Trans-

formation is defined as the uptake and expression of foreign DNA by cells. Bacterial transformation occurs naturally in many species such as *Micrococcus*, *Haemophilus* and *Bacillus*^{1,2}; all these organisms have proteins on their exterior surface whose function is to bind to DNA in their environment and transport it into the cell. However, it is still a rare event for most bacteria to naturally take up DNA from the environment. But by subjecting bacteria to certain artificial conditions, many of them are able to take up free DNA³, and the cells in such state are referred to as competent.

In *E. coli* the competence can be developed by suspending the cells in ice-cold CaCl_2 and then subjecting to a brief heat-shock at 42°C (refs 4, 5). Although *E. coli* has developed into a universal host organism both for molecular cloning and for a diverse set of assays involving cloned genes, the technique of *E. coli* transformation is highly inefficient even using competent cells. The vast majority of DNA molecules added will not enter any cell, and the vast majority of bacterial cells will receive no DNA. Besides Ca^{2+} ions, other frequently used cations include Mg^{2+} , Mn^{2+} , Rb^+ , for competence generation³. However, the exact mechanism by which DNA adsorbs to the *E. coli* cell surface and enters the cell cytosol, and why the transformation is stimulated by these treatments, is still largely obscure.

One proposed hypothesis⁶ is that DNA crosses through the least-barrier path at zones of adhesion, where the outer and inner cell membranes fuse to pores in the cell wall. The zones of adhesion are rich in negatively-charged lipo polysaccharide (LPS) molecules^{7,8} and DNA also being negatively charged, cannot enter the cell easily as the two negative polarities repel each other. A divalent cation such as calcium, is believed to form stable coordination complexes with phosphates, and thus may facilitate the association of the two phosphate-rich structures like DNA and LPS.

The transformation process, being a membrane-bound phenomenon, is most likely to be influenced by the well-

*For correspondence. (e-mail: tbasu@cal3.vsnl.net.in)