ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΙΔΡΥΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΡΕΥΝΑΣ
ΙΝΣΤΙΤΟΥΤΟ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ
ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΔΟΜΙΚΗ ΜΕΛΕΤΗ ΤΗΣ *Pvu*II ΕΝΔΟΝΟΥΚΛΕΑΣΗΣ

ΑΛΕΚΟΣ ΑΘΑΝΑΣΙΑΔΗΣ

ΜΑΡΤΙΟΣ 1995

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΙΔΡΥΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΡΕΥΝΑΣ
ΙΝΣΤΙΤΟΥΤΟ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ
ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΔΟΜΙΚΗ ΜΕΛΕΤΗ ΤΗΣ *Pvu*II ΕΝΔΟΝΟΥΚΛΕΑΣΗΣ

ΑΛΕΚΟΣ ΑΘΑΝΑΣΙΑΔΗΣ

ΜΑΡΤΙΟΣ 1995

Θα ήθελα να αφιερώσω την εργασία αυτή
σε όσους αμέλησα για την πραγματοποίηση της.

Πρόλογος

Η εργασία που περιγράφεται πραγματοποιήθηκε στο σύνολο της στο εργαστήριο Κρυσταλλογραφίας Πρωτεινών, που ανήκει στο τμήμα Βιολογίας του Πανεπιστημίου Κρήτης και στο Ινστιτούτο Μοριακής Βιολογίας και Βιοτεχνολογίας (IMBB), υπό τον Αναπλ. Καθηγητή Μ. Κοκκινίδη. Θα ήθελα να ευχαριστήσω όλους όσους συνετέλεσαν στην δημιουργία και στην λειτουργία του Πανεπιστημίου και του Ινστιτούτου που μου παρείχε την δυνατότητα να πραγματοποιήσω την εργασία αυτή στον Ελληνικό χώρο. Ευχαριστώ θερμά τον καθ. Μ. Κοκκινίδη που αφείδωλα μου παρείχε τα μέσα και την υποστήριξη για την πραγματοποίηση της εργασίας.

Ιδιαίτερα ευχαριστώ τις συνεργάτιδες μου Ντίνα Κοτσιφάκη και Δρ. Μεταξία Βλάση, χωρίς την συμβολή και την συμπαράσταση τους η πραγματοποίηση της εργασίας αυτής θα ήταν αδύνατη για μένα.

Θα ήθελα να ευχαριστήσω επίσης τους ανθρώπους των εργαστηρίων Ενζυμικής Τεχνολογίας για την παροχή των μέσων παραγωγής του υλικού της εργασίας αυτής και Μοριακής Βιολογίας Θηλαστικών για την παραχώρηση του κλώνου της *Pvu*II ενδονουκλεάσης και την θαυμάσια συνεργασία στο επίπεδο της ανάλυσης του.

Πολύτιμη η προσφορά του Dr. Paul Tucker στην διαδικασία συλλογής δεδομένων στις εγκαταστάσεις του EMBL στη Χαϊδελβέργη, καθώς και των Dr. K. Wilson και Dr. Z. Dauter και του συναδέλφου Δ. Παρασκευή για τις μετρήσεις που πραγματοποιήθηκαν στο παράρτημα του EMBL στο DEUTSCHES ELEKTRONEN SYNCHROTRON (DESY) του Αμβούργου.

Σημαντική βοήθεια αποτέλεσαν οι χρηματοδοτήσεις των διεθνών οργανισμών EMBO (European Molecular Biology Organization) και LIP (Large Installation Project της ευρωπαικής ένωσης) για τις απαραίτητες μεταβάσεις στο εξωτερικό.

Πολύτιμες οι συμβουλές του Δρ. Γ. Θηραίου, της Δρ. Μ. Ρίνα και του Δρ. Κ. Πετράτου καθώς και των συναδέλφων μου Δρ. Δ. Καφετζόπουλου, Δρ. Γ. Βλατάκη, Δρ. Θ. Μιχαηλίδη και Γ. Παπανικολάου και ευχαρίστηση κάθε συζήτηση που είχα μαζί τους.

Τις πιό μεγάλες ευχαριστίες μου, στους φίλους μου και στην οικογένεια μου που αποτέλεσαν πηγή δύναμης τα χρόνια αυτά.

Αλέκος Αθανασιάδης

Summary

R.*Pvu*II is a restriction enzyme belonging to the *Pvu*II type II Restriction-Modification system (RMs) of *Proteus vulgaris* species. The recognition sequence of R.*Pvu*II, as of its counterpart methylase M.*Pvu*II, is the CAGCTG palindromic hexanucleotide. R.*Pvu*II cleaves this sequence between the central GC bases leaving blunt ended DNA fragments. The genes of the *Pvu*II RMs have been cloned by Blumenthal et al. and have been shown to be plasmid encoded and located in neighboring DNA regions.

R.*Eco*RI and R.*Eco*RV and recently R.*Bam*HI are the only other type II restriction enzymes for which the 3-dimensional structures are known. Before the completion of the present work and that on R.*Bam*HI despite some similarities of the active sites, the known structures of restriction endonucleases were dissimilar and were using different strategies to attack DNA, leaving this way open questions on the catalytic mechanism and DNA sequence recognition of these enzymes.

In this study we describe the crystallographic determination of the 3D-structure of the R.*Pvu*II enzyme as a free protein, at 2.4Å resolution. Also we compare this structure with the other known structures of restriction enzymes and that of the R.*Pvu*II complexed with cognate DNA.

Initially we determined the sequence of a 2.7 kb fragment of DNA that codes for the *Pvu*II RMs. A 471 bases long open reading frame (ORF) coding for a protein of 157 aminoacids was identified as R.*Pvu*II. The amino acid sequence of R.*Pvu*II protein is unique among the known protein sequences as no significant homology was detected to any other protein sequence. Production of the protein in *E.coli* cells was followed by purification using standard chromatographic techniques resulting to a nearly homogeneous protein sample (>95% pure). Crystallization of the protein by vapor diffusion using the hanging drop technique was successful, employing ammonium sulfate as precipitant at a pH of approximately 5.0.

The crystals obtained were further improved by the application of macro-seeding techniques and measured with CAD4 diffractometer revealing a unit cell of the $P2_12_12$ space group having dimensions a=84.65Å, b=106.46Å, and c=47.0Å. The asymmetric unit of the cell was estimated that contains two monomers.

Data collection for the native protein was carried out using synchrotron X-ray source (DESY) and gave data up to a resolution of 2.3Å. Heavy atom derivative data collection was done either using the same synchrotron X-ray source or a rotating anode source located at EMBL-Heidelberg. Measurement of

the reflections was either carried out using an image plate or a multiwire electronic area detector.

A total of four different isomorphous heavy atom derivatives (Ag,Pr,Yb,Hg) provided useful phasing information. For the Hg derivative a second independent data set was collected and used. Using five isomorphous data sets and four sets of anomalous data, phases with a figure of merit of 0.59 up to 2.8Å resolution were calculated. These phases were further improved applying solvent flattening and molecular averaging techniques to the resulting electron density.

MIR maps and density modified maps were used for modeling the R.*Pvu*II protein and resulted to a first model in which 14 loop residues were missing. Repeated cycles of crystallographic refinement of the model using simulated annealing technique, conventional refinement and subsequent modeling gave rise to a final model of the protein with an R-factor of 18.4%, missing two aminoacids of the total 156 of the mature protein.

The enzyme has a mixed α/β structure with an extended twisted β-sheet forming the core of the protein. Two subdomains can be easily recognized in the protein structure. The first consists of the 35 N-terminal amino acids and provides the majority of the dimmerization interactions between the two monomers. Comparison of the R.*Pvu*II 3D-structure with the R.*Eco*RV structure revealed topological similarities between the secondary structure elements of the two proteins. This similarity allowed us to assign to the second subdomain (consisting of the rest of the molecule) the catalytic and DNA recognition functions (This assignment was later confirmed from the structure of R.*Pvu*II complexed with cognate DNA {Cheng et al., 1994}). The topological similarity between endonucleases which are unrelated in their primary structures lead us to suggest the classification of restriction enzymes based on the position of the scissile phosphate bond and the structural restrictions this implies.

Compared with the DNA binding cleft of R.*Eco*RV or R.*Pvu*II complexed with cognate DNA, the free enzyme exhibits a considerably more wide cleft which suggests a flexibility of the enzyme leading to drastic conformational changes upon interaction with DNA. The fact that DNA-sequence specific interactions shown in the structure of the R.*Pvu*II cognate complex are not feasible in the free enzyme conformation, suggests that recognition interactions probably occur in distinct stages which are coupled with changing protein conformations.

Περίληψη

Η R.*Pvu*II είναι ένα περιοριστικό ένζυμο που ανήκει στο *Pvu*II, τύπου II, σύστημα περιορισμού και τροποποίησης (RMs) του DNA, από το βακτήριο *Proteus vulgaris*. Η αλληλουχία αναγνώρισης για τη R.*Pvu*II πρωτεΐνη, όπως και για το ταίρι της την M.*Pvu*II (*Pvu*II μεθυλάση), είναι η CAGCTG. Η R.*Pvu*II τέμνει την αλληλουχία μεταξύ των δύο κεντρικών βάσεων δημιουργώντας κομμάτια DNA με "τυφλά" άκρα. Τα γονίδια που κωδικοποιούν το *Pvu*II RMs έχουν κλωνοποιηθεί από τους Blumenthal et al. οι οποίοι έδειξαν παράλληλα, ότι αυτά βρίσκονται σε γειτονικές θέσεις στον πλασμιδιακό φορέα pPvuI.

Η R.*Eco*RI και η R.*Eco*RV και πρόσφατα η R.*Bam*HI είναι οι μόνες άλλες τύπου II ενδονουκλεάσες για τις οποίες είναι γνωστή η τρισδιάστατη δομή. Πρίν την ολοκλήρωση της παρούσης εργασίας όπως και αυτής της R.*Bam*HI, παρά τις ομοιότητες στο ενεργό κέντρο τους, οι γνωστές δομές ενδονουκλεασών εμφανιζόταν δομικά ανόμοιες και ακολουθούσαν διαφορετικές στρατηγικές για την πρόσδεση στο μόριο του DNA, αφήνοντας ανοικτά ερωτήματα αναφορικά με τον καταλυτικό μηχανισμό και την αναγνώριση του DNA από τα ένζυμα αυτά.

Στην εργασία αυτή περιγράφεται ο προσδιορισμός της τρισδιάστατης δομής της *Pvu*II ενδονουκλεάσης στην μορφή της ελεύθερης πρωτεΐνης, σε διακριτικότητα 2.4Å. Επίσης, παρουσιάζεται η σύγκριση της δομής της R.*Pvu*II με τις προηγούμενα γνωστές δομές περιοριστικών ενζύμων καθώς και με αυτήν του συμπλόκου της R.*Pvu*II με την συγγενή αλληλουχία DNA.

Αρχικά προσδιορίστηκε η νουκλεοτιδική αλληλουχία ενός τμήματος DNA μεγέθους 2.7kb που κωδικοποιεί το *Pvu*II RMs. Ενα ανοιχτό πλαίσιο διαβάσματος, που κωδικοποιεί μία πρωτεΐνη 157 καταλοίπων, αναγνωρίστηκε σαν η *Pvu*II ενδονουκλεάση. Η πρωτεϊνική αλληλουχία της *Pvu*II ενδονουκλεάσης είναι μοναδική μεταξύ των γνωστών πρωτεϊνικών αλληλουχιών καθώς δεν στάθηκε δυνατόν να ανιχνευτεί καμμία σημαντική ομοιότητα. Η παραγωγή της πρωτεΐνης σε κύτταρα *E.coli* ακολουθήθηκε από καθαρισμό με την χρήση συμβατικής χρωματογραφίας έχοντας ως αποτέλεσμα ομογενώς καθαρό πρωτεϊνικό δείγμα (>95% καθαρή πρωτεΐνη). Κρυστάλλωση της πρωτεΐνης με την μέθοδο "hanging drop" υπήρξε επιτυχής με την χρήση θειικού αμμωνίου ως παράγοντα κατακρήμνισης της πρωτεΐνης, σε pH 5.0.

Οι κρύσταλλοι που προέκυψαν βελτιώθηκαν περαιτέρω εφαρμόζοντας την τεχνική της μάκρο-σποράς. Κρυσταλλογραφικά δεδομένα με χρήση ακτίνων X συλλέχθηκαν με ένα CAD4 περιθλασίμετρο αποκαλύπτοντας μία στοιχειώδη κυψελίδα που ανήκει στην χωροομάδα συμμετρίας $P2_12_12$ με διαστάσεις a=84.65Å,

b=106.46Å, και c=47.00Å ενώ υπολογίστηκε ότι η ασύμμετρη μονάδα της κυψελίδας περιέχει δύο μονομερή.

Η συλλογή δεδομένων για την φυσική πρωτεΐνη πραγματοποιήθηκε με την χρήση ακτινοβολίας σύγχροτρον και έδωσε δεδομένα, διακριτικότητας, μέχρι 2.3Å. Για τα παράγωγα βαρειών ατόμων η συλλογή δεδομένων έγινε με την χρήση της ίδιας πηγής ακτινοβολίας σύγχροτρον ή με πηγή περιστρεφόμενης ανόδου στο EMBL στην Χαϊδελβέργη. Οι μετρήσεις των ανακλάσεων πραγματοποιήθηκαν με δίσκο ειδώλου ή ανιχνευτή περιοχής.

Συνολικά 4 διαφορετικά παράγωγα βαρειών ατόμων (Ag,Pr,Yb,Hg) αποδείχθηκαν χρήσιμα για τον προσδιορισμό των φάσεων της πρωτεΐνης. Για το παράγωγο του Hg συλλέχθηκε και χρησιμοποιήθηκε ένα δεύτερο ανεξάρτητο σύνολο δεδομένων. Χρησιμοποιώντας 5 σύνολα δεδομένων ισόμορφων παραγώγων και δεδομένα ανώμαλου σκεδασμού για 4 από αυτά, προσδιορίστηκαν φάσεις της πρωτεΐνης με δείκτη αξιοπιστίας (FOM) 0.59. Οι φάσεις αυτές βελτιώθηκαν περαιτέρω με την εφαρμογή των τεχνικών εξομάλυνσης διαλύτη (solvent flattening) και μοριακής εξισορρόπισης (molecular averaging) στη προκύπτουσα ηλεκτρονιακή πυκνότητα.

MIR χάρτες και χάρτες με τροποποιημένη ηλεκτρονιακή πυκνότητα χρησιμοποιήθηκαν για την κατασκευή του πρώτου μοντέλου της πρωτεΐνης από το οποίο απουσίαζαν 14 αμινοξικά κατάλοιπα. Βελτιστοποίηση του μοντέλου χρησιμοποιώντας την τεχνική "simulated annealing" καθώς και συμβατικές τεχνικές βελτιστοποίησης παράλληλα με μοντελοποίηση, οδήγησε στο τελικό μοντέλο της πρωτεΐνης με παράγοντα αξιοπιστίας R 18.4% για 154 από τα 156 αμινοξέα της ώριμης πρωτεΐνης.

Η πρωτεΐνη έχει μικτή α/β δομή με μία εκτεταμένη και περιεστραμένη β-πτυχωτή επιφάνεια να αποτελεί τον πυρήνα της δομής του ενζύμου. Δύο υποπεριοχές μπορούν εύκολα να διακριθούν στην δομή. Η πρώτη αποτελούμενη από τα 35 αμινοτελικά αμινοξικά κατάλοιπα της πρωτεΐνης σχηματίζει το μεγαλύτερο μέρος από τις επαφές μεταξύ των δύο μονομερών και χαρακτηρίζεται ως υποπεριοχή διμερισμού. Σύγκριση της δομής της R.*Pvu*II με αυτήν της R.*Eco*RV αποκάλυψε εκτεταμένες τοπολογικές ομοιότητες μεταξύ των στοιχείων δευτεροταγούς δομής των δύο πρωτεϊνών. Αυτή η τοπολογική ομοιότητα μας επέτρεψε την απόδοση, στην δεύτερη υποπεριοχή που περιλαμβάνει το υπόλοιπο της πρωτεΐνης, των λειτουργιών της κατάλυσης και της αναγνώρισης του DNA. Η τοπολογική ομοιότητα μεταξύ δύο πρωτεϊνών που δεν σχετίζονται σε επίπεδο πρωτοταγούς δομής μας οδήγησε να προτείνουμε την ομαδοποίηση των περιοριστικών ενζύμων με βάση την θέση του δεσμού που διασπάται στην αλυσίδα του DNA και τους δομικούς περιορισμούς που τίθενται από αυτήν.

Συγκρίνοντας με το άνοιγμα πρόσδεσης του DNA της R.*Eco*RV ή της R.*Pvu*II προσδεδεμένης σε συγγενή αλληλουχία του DNA, η ελεύθερη πρωτεΐνη επιδεικνύει ένα σημαντικά πλατύτερο άνοιγμα, το οποίο υποδεικνύει την ευελιξία της δομής του ενζύμου που έχει ως αποτέλεσμα δραστικές αλλαγές στην διαμόρφωση κατά την πρόσδεση στο DNA. Επιπλέον το γεγονός ότι οι ειδικές αλληλεπιδράσεις που παρουσιάζονται στη δομή του συμπλόκου με την συγγενή αλληλουχία του DNA δεν είναι εφικτές από την ελεύθερη μορφή της πρωτεΐνης οδηγεί στην υπόθεση ότι οι αλληλεπιδράσεις αναγνώρισης του DNA είναι δυνατόν να συμβαίνουν σε χρονικά διακριτά στάδια συνδυασμένες με διαφορετικές διαμορφώσεις της πρωτεΐνης.

Περιεχόμενα

Πρόλογος.....1
Summary.....2
Περίληψη.....4
Περιεχόμενα.....7

Α. Εισαγωγή.....10

Α 1. Συστήματα περιορισμού και τροποποίησης του DNA.....11
Α 2. Τύπου ΙΙ συστήματα περιορισμού και τροποποίησης του DNA.....12
Α 3. Δομική πληροφορία για τα τύπου ΙΙ περιοριστικά ένζυμα.....14
Α 4. Το PvuII RM σύστημα.....22
Α 5. Σκοπός της περιγραφόμενης εργασίας.....22

Β. Αποτελέσματα και Συζήτηση.....24

Β 1. Προσδιορισμός και ανάλυση της νουκλεοτιδικής αλληλουχίας του PvuII συστήματος.....25
Β 1.1 Στρατηγική του προσδιορισμού της νουκλεοτιδικής αλληλουχίας.....25
Β 1.2. Ανάλυση της νουκλεοτιδικής αλληλουχίας της R.PvuII.....26
Β 2. Πρόβλεψη της αλληλουχίας της "star" (*) ενεργότητας της PvuII ενδονουκλεάσης σε συνθήκες υπερπέψης.....31
Β 3. Απομόνωση και καθαρισμός του ενζύμου.....34
Β 4. Κρυστάλλωση της PvuII ενδονουκλεάσης.....38
Β 5. Συλλογή κρυσταλλογραφικών δεδομένων.....42
Β 5.1. Αρχικός κρυσταλλογραφικός χαρακτηρισμός των κρυστάλλων της ενδονουκλεάσης PvuII.....42
Β 5.2. Συλλογή δεδομένων για την φυσική πρωτεΐνη.....43
Β 5.3. Προετοιμασία και συλλογή δεδομένων των παραγώγων βαρειών ατόμων των κρυστάλλων της R.PvuII.....46
Β 6. Προκαταρκτική ανάλυση των δεδομένων της φυσικής πρωτεΐνης: Ο τάξεως-2 άξονας συμμετρίας.....50
Β 7. Πολλαπλή ισόμορφη αντικατάσταση. (Multiple Isomorphous Replacement (MIR) phasing).....53
Β 7.1 Αναγωγή δεδομένων των παραγώγων με βαριά άτομα στην κλίμακα των δεδομένων των κρυστάλλων της φυσικής πρωτεΐνης. Στατιστική ανάλυση των

διαφορών των δεδομένων παραγώγων με βαριά άτομα - φυσικής πρωτεΐνης.....53
Β 7.2. Προσδιορισμός και βελτιστοποίηση των θέσεων βαρειών ατόμων.....56
Β 8. Κατασκευή του μοντέλου της R.PvuII.....75
Β 8.1. Ο σκελετός και το αρχικό μοντέλο.....75
Β 8.2. Βελτιστοποίηση του μοντέλου της R.PvuII.....76
Β 8.3. Ανάλυση της ποιότητας του μοντέλου.....80
Β 9. Ανάλυση του μοντέλου της R.PvuII.....85
Β 9.1. Χαρακτηρισμός των στοιχείων της δευτεροταγούς δομής.....85
Β 9.2. Η δομή του μορίου στις τρείς διαστάσεις.....85
Β 9.3. Πακετάρισμα των μορίων στην κρυσταλλική δομή.....92
Β 9.4. Θέσεις της πρόσδεσης των βαρειών ατόμων στην πρωτεΐνη.....95
Β 10. Συγκριτική μελέτη της δομής της R.PvuII.....98
Β 10.1. Συσχέτιση των δομών της R.PvuII και R.EcoRV.....98
Β10.2. Σύγκριση της ελεύθερης R.PvuII με την πρωτεΐνη σε σύμπλοκο με συγγενές DNA.....107

Γ. Συμπεράσματα και προοπτικές.....115

Δ. Υλικα και Μέθοδοι.....122

Δ 1. Υλικά.....123
Δ 2. Απομόνωση και ανάλυση νουκλεικών οξεών.....124
Δ 2.1. Απομόνωση πλασμιδιακού DNA.....124
Δ 2.2. Ανάλυση τμημάτων DNA σε πήκτωμα αγαρόζης.....124
Δ 2.3. Προσδιορισμός νουκλεοτιδικής αλληλουχίας μορίων DNA.....124
Δ 3. Παραγωγή, απομόνωση και χαρακτηρισμός της πρωτεΐνης.....125
Δ 3.1. Μεγάλης κλίμακας καλλιέργεια βακτηρίων για την παραγωγή πρωτεΐνης.....125
Δ 3.2. Λύση πάστας βακτηριακών κυττάρων.....126
Δ 3.3. Προετοιμασία υλικού για την κατασκευή στήλης φωσφοκυτταρίνης (PC).....127
Δ 3.4. Μέθοδος καθαρισμού της PvuII ενδονουκλεάσης.....128
Δ 3.5. Αντίδραση πέψης DNA από κλάσματα του καθαρισμού.....129
Δ 3.6. Χρώση πρωτεϊνών με νιτρικό άργυρο, ύστερα από ανάλυση τους σε πηκτώματα ακρυλαμίδης.....129
Δ 3.7. Ποσοτικός προσδιορισμός πρωτεϊνών.....130
Δ 3.8. Αντίδραση υπερπέψης DNA (Overdigest).....130
Δ 4. Κρυστάλλωση και χειρισμός των κρυστάλλων της R.PvuII.....131
Δ 4.1. Κρυστάλλωση της R.PvuII με την "hanging drop" μέθοδο.....131

Δ 4.2. Μάκρο-σπορά στους κρυστάλλους της R.PvuII. (Macroseeding) 131
Δ 4.3. Προετοιμασία παραγώγων βαρειών ατόμων..... 132
Δ 4.4. Προετοιμασία κρυστάλλων για έκθεση σε ακτίνες X..... 132
Δ 5. Προσδιορισμός της δομής της R.PvuII..... 134
Δ 5.1. Επεξεργασία δεδομένων από δίσκο ειδώλου..... 134
Δ 5.2. Συνθέσεις Patterson και Fourier..... 135
Δ 5.3. Εύρεση θέσεων των βαρειών ατόμων από συνθέσεις Patterson..... 135
Δ 5.4. Βελτιστοποίηση των θέσεων των βαρειών ατόμων..... 136
Δ 5.5. Προσδιορισμός των φάσεων της πρωτεΐνης..... 136
Δ 5.6. Απεικόνηση χαρτών ηλεκτρονιακής πυκνότητας καθώς και μοντέλων της πρωτεΐνης..... 136
Δ 5.7. Τροποποίηση ηλεκτρονιακής πυκνότητας..... 137
Δ 5.8. Κρυσταλλογραφική βελτιστοποίηση του μοντέλου της R.PvuII..... 138
Δ 6. Χειρισμός των μοντέλων δομής..... 138
Δ 6.1. Ανάλυση του μοντέλου της R.PvuII..... 138
Δ 6.2. Συγκρίσεις μεταξύ μοντέλων των R.PvuII και R.EcoRV, R.PvuII σε σύμπλοκο με DNA..... 139
Προσάρτημα I 140
Προσάρτημα II 141
Προσάρτημα III 142
Απόδοση στην Ελληνική γλώσσα μή δόκιμων όρων της διεθνούς ορολογίας..... 143
ΒΙΒΛΙΟΓΡΑΦΙΑ..... 145

Α. Εισαγωγή

Α 1. Συστήματα περιορισμού και τροποποίησης του DNA.

Η τροποποίηση του DNA στους προκαρυοτικούς οργανισμούς είναι συνυφασμένη κατά κύριο λόγο με τα συστήματα περιορισμού και τροποποίησης (Restriction-Modification systems : RMs) {Nathans & Smith, 1975}. Ενα RMs κατά κανόνα χαρακτηρίζεται από δύο ενεργότητες, μία ικανή να μεταφέρει ένα μεθύλιο από το μόριο S-αδενοσυλ-L-μεθειονίνη (SAM) σε μία βάση του DNA (μεθυλάση) και μία ικανή να κόβει από ένα φωσφοδιεστερικό δεσμό στις δύο αλυσίδες αμεθυλίωτου DNA ανάμεσα σε συγκεκριμένες βάσεις (ενδονουκλεάση) {Moldrich & Roberts, 1982}, {Wilson & Murray, 1991}. Οι δύο ενεργότητες ενός RMs είναι ειδικές, ενεργούν δηλαδή πάνω σε συγκεκριμένη και κοινή αλληλουχία DNA. Η προστασία του ενδογενούς DNA από την ενεργότητα του περιοριστικού ενζύμου καθιστά αναγκαία την συμπαρουσία της ενεργότητας μεθυλίωσης του DNA σε οργανισμούς που φέρουν RMs.

Η εκτεταμένη αναζήτηση RMs για λόγους που σχετίζονται με την χρησιμότητα τους στις διαδικασίες κλωνισμού {Brooks, 1987} έχει οδηγήσει στην ανακάλυψη περισσότερων από 2400 τέτοιων συστημάτων τόσο σε βακτήρια όσο και σε βακτηριακούς ιούς (φάγους) και ιούς των μονοκύτταρων φυκών {Roberts & Halford, 1993}, {Roberts & Macelis, 1992}. Το ένα τέταρτο περίπου των οργανισμών πού έχουν εξετασθεί για την παρουσία RMs έχουν ένα η περισσότερα τέτοια συστήματα ενώ περίπου δύο νέα RMs προστίθενται στον κατάλογο σε καθημερινή βάση και η πληροφορία πάνω στο θέμα αριθμούσε 2380 δημοσιευμένα άρθρα μέχρι το τέλος του 1992 {Roberts, 1992}.

Ο πιθανός βιολογικός ρόλος των συστημάτων αυτών είναι η προστασία των βακτηριακών κυττάρων από την μολυσματική δράση ξένου DNA και ειδικότερα DNA φάγου {Kruger & Bickle, 1983}. Αυτό επιτυγχάνεται με την πέψη του νεοεισερχόμενου στο κύτταρο DNA από την περιοριστική δράση του συστήματος με ταυτόχρονη προστασία μέσω μεθυλίωσης του ενδογενούς DNA.

Τα RMs διαφοροποιούνται σε μία σειρά κατηγοριών με βάση τις ανάγκες τους σε συμπαράγοντες, την συμμετρία της αλληλουχίας αναγνώρισης, τον αριθμό των πολυπεπτιδικών αλυσίδων καθώς και τα χαρακτηριστικά κοπής του DNA (πίνακας 1) {Zabau & Roberts, 1979}.

Σε όλους τους τύπους RMs κοινή είναι η αναγκαιότητα της παρουσίας Mg^{2+} για την πέψη του DNA και της S-αδενοσυλιομένης L-μεθειονίνης (SAM) ως δότη μεθυλίου για την αντίδραση μεθυλίωσης του DNA.

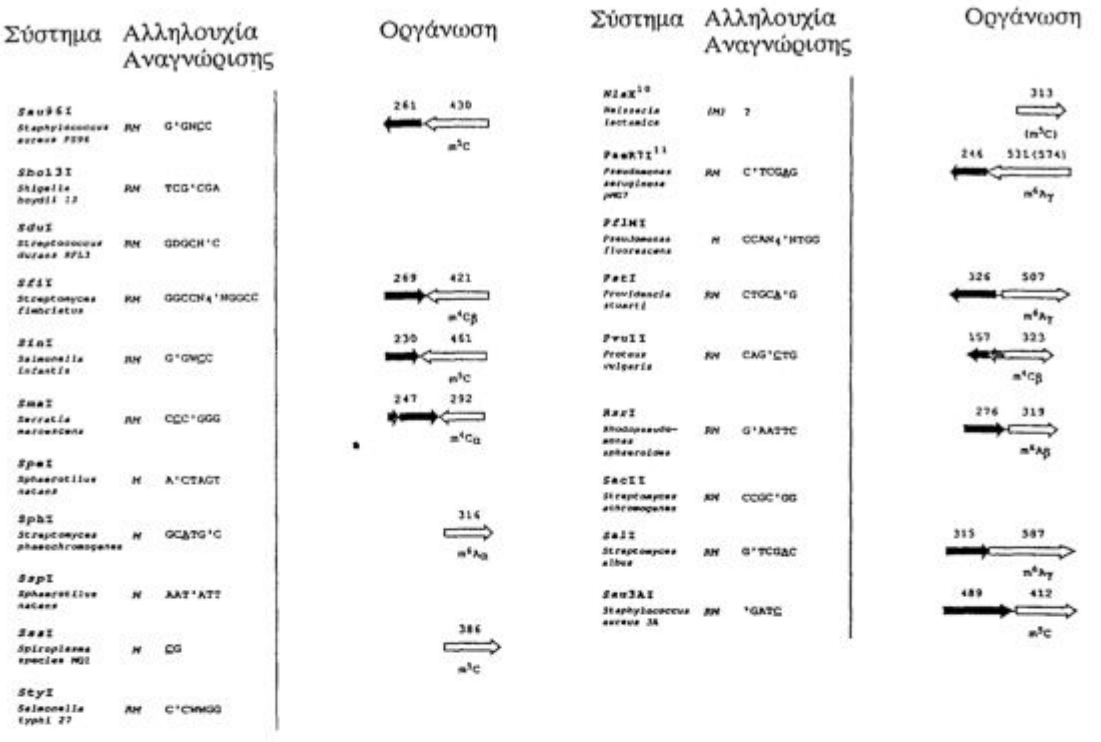
Πίνακας 1. Κατηγορίες των RMs {Zabau & Roberts, 1979}.[1]

Τύπου I.	Τρείς πολυπεπτιδικές αλυσίδες R, M και S συνθέτουν ένα ένζυμο ικανό να μεθυλιώνει και να πέπτει το DNA. Απαραίτητοι συμπαράγοντες για την πέψη του DNA είναι το SAM και το ATP. Το σημείο τομής του DNA απέχει σημαντικά από το σημείο αναγνώρισης της αλληλουχίας ενώ η αλληλουχία είναι μη συμμετρική και διαμερισμένη. Η ικανότητα αναγνώρισης της αλληλουχίας του DNA εντοπίζεται στην υπομονάδα S και προσδιορίζει την ειδικότητα τόσο για την τροποποίηση όσο και για την τομή του DNA.
Τύπου II.	Τα τύπου II RMs είναι τα απλούστερα και πιό κοινά συστήματα. Οι δύο ενεργότητες εντοπίζονται σε ξεχωριστές πρωτεΐνες και αναγνωρίζουν παλίνδρομες αλληλουχίες. Το μέγεθος της αλληλουχίας αναγνώρισης κυμαίνεται από 4 έως 10 βάσεις. Η τομή του DNA γίνεται συμμετρικά και πάντα στο εσωτερικό της αλληλουχίας αναγνώρισης. Στις περιπτώσεις που έχουν μελετηθεί τα τύπου II περιοριστικά ένζυμα λειτουργούν ώς ομοδιμερή ενώ οι μεθυλάσες ως μονομερή.
Τύπου IIs.	Ασύμμετρες αλληλουχίες αναγνώρισης μεγέθους 5-7 βάσεων χαρακτηρίζουν τα τύπου IIs συστήματα. Η τομή του DNA γίνεται σε καθορισμένη απόσταση, μέχρι 20 βάσεις από το σημείο αναγνώρισης. Στην περίπτωση αυτή οι ενδονουκλεάσες πιθανότατα δρούν ως μονομερή και είναι σημαντικά μεγαλύτερες από τα αντίστοιχα ένζυμα των τύπου II συστημάτων.
Τύπου III.	Στα τύπου III συστήματα το περιοριστικό ένζυμο και η μεθυλάση σχηματίζουν σύμπλοκο ικανό να μεθυλιώνει και να πέπτει το DNA. Η μεθυλάση είναι ικανή να μεθυλιώσει και απουσία του περιοριστικού. Η πέψη απαιτεί ATP και ενισχύεται παρουσία SAM. Η αλληλουχία αναγνώρισης είναι ασύμμετρη και μεγέθους 5-6 βάσεων ενώ η τομή του DNA γίνεται σε απόσταση 25 περίπου βάσεων από το σημείο αναγνώρισης.

[1] Σε μερικές περιπτώσεις έχουν βρεθεί συστήματα που δεν μπορούν να ενταχθούν στις κατηγορίες του πίνακα.

Α 2. Τύπου II συστήματα περιορισμού και τροποποίησης του DNA.

Τα τύπου II συστήματα περιορισμού και τροποποίησης (RMs II) αποτελούν την πιό μελετημένη κατηγορία RMs σε μοριακό, βιοχημικό και δομικό επίπεδο. Για περισσότερα από 60 RMs II έχει προσδιοριστεί η νουκλεοτιδική αλληλουχία και έχει αναλυθεί η γονιδιακή οργάνωση τους, ενώ περισσότερα από 100 έχουν κλωνοποιηθεί. Κοινό χαρακτηριστικό των συστημάτων αυτών αποτελεί η οργάνωση των γονιδίων τους (εικόνα 1) τα οποία εδράζονται σε γειτονικές θέσεις στο DNA {Wilson, 1988, Wilson, 1991}. Τα γονίδια του περιοριστικού και της μεθυλάσης σε άλλες περιπτώσεις μεταγράφονται στην ίδια και άλλοτε σε αντίθετη φορά. Η μεταφορά των RMs II από πλασμιδιακούς φορείς είναι πολύ συχνή αν και η ενσωμάτωση τους στο βακτηριακό χρωμόσωμα συναντιέται επίσης. Το περιοριστικό ένζυμο είναι κατά κανόνα μικρότερο από την αντίστοιχη μεθυλάση και σε μερικές



Εικόνα 1. Αντιπροσωπευτική γονιδιακή οργάνωση των τύπου II RMs. {Wilson, 1991}. *Με βέλη τα γονίδια των ενδονουκλεασών (μαύρα), των μεθυλασών (άσπρα) και των ρυθμιστικών πρωτεϊνών (σκιασμένα). Πάνω από τα βέλη σημειώνεται ο αριθμός αμινοξικών καταλοίπων της πρωτεΐνης.*

περιπτώσεις (*Bam*HI, *Bsu*RI, *Eco*RV, *Eco*72I, *Mun*I, *Mva*I, *Pvu*II και *Sma*I) ρυθμίζεται σε μεταγραφικό επίπεδο από μία πρωτεΐνη μεγέθους 60-80 αμινοξέων της οποίας το γονίδιο συνοδεύει αυτό του περιοριστικού και της μεθυλάσης {Tao, Bourne and Blumenthal, 1991}. Ο ακριβής μηχανισμός της ρύθμισης αυτής παραμένει άγνωστος μέχρι στιγμής.

Συγκρίσεις των γνωστών αμινοξικών αλληλουχιών των ενδονουκλεασών τύπου II με τις αντίστοιχες μεθυλάσες χαρακτηρίζονται από την απουσία οποιασδήποτε σημαντικής ομοιότητας {Chandrasegaran & Smith, 1988}, παρά το γεγονός της αναγνώρισης της ίδιας αλληλουχίας DNA. Η ίδια απουσία ομολογιών εμφανίζεται και ανάμεσα στις αλληλουχίες των ενδονουκλεασών, ακόμα και στις περιπτώσεις ετεροσχιζομερών, αυτών δηλαδή που αναγνωρίζουν και πέπτουν την ίδια αλληλουχία σε διαφορετικά σημεία {Chandrasegaran & Smith, 1988}. Ομολογίες εμφανίζονται, σε μερικές περιπτώσεις, μόνο μεταξύ ομοσχιζομερών, ενζύμων που η αλληλουχία αναγνώρισης αλλά και το σημείο τομής ταυτίζονται {Stephenson et al., 1989}. Αυτές είναι οι μόνες περιπτώσεις στις οποίες η παρουσία κοινού εξελικτικά προγόνου μπορεί να υποτεθεί με δεδομένα από τις αναλύσεις

των πρωτεϊνικών αλληλουχιών. Στο σύνολο τους τα περιοριστικά αντιπροσωπεύουν ένα εξελικτικό αίνιγμα καθώς κάθε αλληλουχία είναι μοναδική {Wilson, 1991} και δεν εμφανίζει κανένα από τα γνωστά πρότυπα αλληλουχιών πρωτεϊνών που δεσμεύονται στο DNA. Το πρόβλημα περιπλέκεται περισσότερο καθώς οι πρωτεϊνικοί παράγοντες που ρυθμίζουν την μεταγραφή των γονιδιών των ενδονουκλεασών δείχνουν σημαντικές ομολογίες ανάμεσα τους {Tao, Bourne and Blumenthal, 1991}.

Αντίθετα με τα περιοριστικά ένζυμα, οι μεθυλάσες χαρακτηρίζονται, σε επίπεδο αλληλουχιών, από εκτεταμένες ομοιότητες μεταξύ τους. Οι ενενήντα γνωστές αλληλουχίες μεθυλασών ομαδοποιούνται σε επτά διαφορετικές κατηγορίες {Klimasauskas et al., 1989}. Η ομαδοποίηση των μεθυλασών με βάση χαρακτηριστικά πρότυπα στην αλληλουχία τους συμπίπτει με την ομαδοποίηση με βάση βιοχημικά τους χαρακτηριστικά όπως το άτομο της βάσης που δέχεται το μεθύλιο ή την βάση την οποία μεθυλιώνουν {Smith & Kelly, 1984}. Ετσι στην περίπτωση των ενζύμων που δημιουργούν 5-μεθυλοκυτοσίνη (m^5C μεθυλάσες) περισσότερα από δέκα κοινά πρότυπα έχουν ταυτοποιηθεί στις αλληλουχίες τους {Posfai et al., 1989}, {Klimasauskas, Nelson and Roberts, 1991}. Ο δομικός ρόλος των συντηρημένων προτύπων των μεθυλασών αποκαλύπτεται στην πρόσφατη δομική ανάλυση της *Hha*I μεθυλάσης {Cheng et al., 1994}.

Α 3. Δομική πληροφορία για τα τύπου ΙΙ περιοριστικά ένζυμα.

Ιδιαίτερο ενδιαφέρον έχει υπάρξει τα τελευταία χρόνια για την δομή των τύπου ΙΙ ενδονουκλεασών με αποτέλεσμα συμπεριλαμβανομένης της R.*Pvu*II να είναι κρυσταλλογραφικά προσδιορισμένες δομές τέσσαρων περιοριστικών ενζύμων από τις οποίες τρείς επιλύθηκαν τα τελευταία δύο χρόνια.

Κυριότεροι λόγοι για το ενδιαφέρον αυτό είναι ότι α) Κανένα από τα γνωστά δομικά πρότυπα πρωτεϊνών που δένονται σε DNA (HLH {Ellenberger, 1994}, Zinc fingers {Schmiedeskamp & Klevit, 1994}, Leucine zippers {Ellenberger, 1994}, HTH {Harrison & Aggarwal, 1990}) δεν φαίνεται να υπακούεται από τα περιοριστικά ένζυμα {Freemont, 1991}. β) Από τα περιοριστικά ένζυμα που έχουν ανακαλυφθεί μέχρι σήμερα αναγνωρίζονται 188 διαφορετικές αλληλουχίες DNA, δημιουργώντας ένα πλούσιο δείγμα για την μελέτη των ειδικών αλληλεπιδράσεων πρωτεϊνών-DNA. γ) Η έλλειψη ομολογιών μεταξύ των περιοριστικών ενζύμων καθιστά πιθανό το γεγονός ότι περισσότεροι του ενός μηχανισμοί αναγνώρισης του DNA έχουν εξελικτικά επιλεχθεί για τον σκοπό αυτό. δ) Η ανάλυση της δομής των μορίων αυτών πιθανόν να δώσει στοιχεία για την εξέλιξη των RMs καθώς και για την εξελικτική σχέση ενδονουκλεασών-

μεθυλασών. ε) Η γνώση της δομής των περιοριστικών ενζύμων πιθανά θα επιτρέψει την σχεδίαση νέων μορίων με διαφορετικές ιδιότητες τόσο καταλυτικές όσο και σε ότι αφορά την ειδικότητα τους, γεγονός με ιδιαίτερο βιοτεχνολογικό ενδιαφέρον.

Την στιγμή που γράφεται η παρούσα εργασία είναι γνωστές οι δομές: της *Eco*RV ενδονουκλεάσης σε μορφή συμπλόκου με συγγενές (cognate) και μή συγγενές (non-cognate) DNA, καθώς και αυτή τής ελεύθερης πρωτεΐνης {Winkler et al., 1993}. Της R.*Eco*RI σε σύμπλοκο με συγγενές DNA {Kim et al., 1990}. Της ελεύθερης πρωτεΐνης της R.*Bam*HI {Newman et al.,1994a, Newman et al.,1994b} και τέλος της R.*Pvu*II σε σύμπλοκο με συγγενές DNA {Cheng et al., 1994} και στην ελεύθερη μορφή της {Athanasiadis et al., 1994}.

Τα ερωτήματα για τα οποία καλείται να δώσει στοιχεία ο συνδυασμός δομικής και λειτουργικής μελέτης των ενδονουκλεασών αφορούν :

α) Το συνολικό δίπλωμα των ενδονουκλεασών. Υπάρχουν κοινά δομικά χαρακτηριστικά ανάμεσα στα περιοριστικά ένζυμα; Ποιά η εξελικτική σχέση μεταξύ των περιοριστικών ενζύμων και ποιά η σχέση τους με άλλες κατηγορίες πρωτεϊνών που προσδένονται στο DNA {Halazonatis & Kandil, 1992};

β) Το μηχανισμό πρόσδεσης στο μόριο του DNA. Πως το ένζυμο αντιλαμβάνεται την παρουσία του DNA ; Πως γίνεται η αρχική πρόσδεση σε τυχαία αλληλουχία DNA;

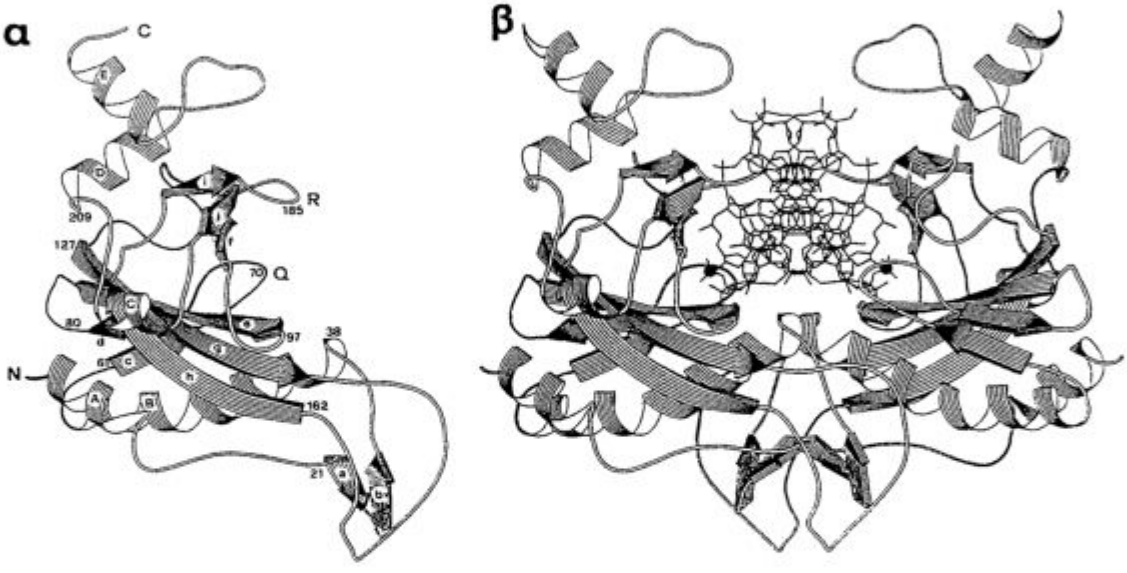
γ) Τον μηχανισμό κίνησης της πρωτεΐνης πάνω στο μόριο του DNA {Taylor et al., 1991}.

δ) Την ανίχνευση του DNA για την αλληλουχία αναγνώρισης. Πως η πρωτεΐνη αναζήτα τον στόχο της και πώς επιτυγχάνεται η μεγάλη ειδικότητα των περιοριστικών ενζύμων (αναγνώριση ενός στα 4096 πιθανά εξανουκλεοτίδια) {Bennet & Halford, 1989};

ε) Τον μηχανισμό υδρόλυσης του DNA. Ποιές αλλαγές κατά την αναγνώριση της αλληλουχίας στόχου ενεργοποιούν τον μηχανισμό διάσπασης του φωσφοδιεστερικού δεσμού; Ποιά η δομή του καταλυτικού κέντρου των περιοριστικών ενζύμων; (ειναι κοινό για όλα τα περιοριστικά;) Ποιός ο ρόλος του Mg^{2+} {Thielking et al., 1991, Thielking at al., 1992};

στ) Ποιός ο ρόλος της δομής του μορίου του DNA στην όλη διαδικασία {Winkler, 1992};

Στην συνέχεια παρουσιάζονται τα δεδομένα που προέκυψαν από την δομική ανάλυση των R.*Eco*RV και R.*Eco*RI που προϋπήρχαν της ολοκλήρωσης της παρούσης εργασίας. Η δομές των R.*Bam*HI καθώς και του συμπλόκου της R.*Pvu*II με συγγενές DNA προέκυψαν στην διάρκεια της συγγραφής της παρούσης και τα μεν αποτελέσματα της R.*Bam*HI παρουσιάζονται συνοπτικά παρακάτω ενώ σε ξεχωριστό κεφάλαιο των αποτελεσμάτων γίνεται η παρουσίαση του συμπλόκου της R.*Pvu*II ταυτόχρονα με την σύγκριση του με την ελεύθερη μορφή της πρωτεΐνης (παρούσα εργασία).



Εικόνα 2. Απεικόνιση της τρισδιάστατης δομής της R.EcoRV. *α) Το μονομερές της ελεύθερης μορφής της R.EcoRV όπου παρουσιάζεται η ονοματολογία των δευτεροταγών στοιχείων της δομής του μορίου. Q και R οι δύο βρόγχοι της πρωτεΐνης που μετέχουν σε αλληλεπιδράσεις με τις βάσεις του DNA. β) Το διμερές της R.EcoRV, στο κέντρο διακρίνεται το μόριο του DNA. Ο άξονας του DNA είναι κάθετος στο επίπεδο του σχήματος ενώ στο μόριο του DNA οι διασπάσιμες φωσφοδιεστερικές ομάδες ως μαύροι κύκλοι.{Winkler et al., 1993}*

Γενικά χαρακτηριστικά των δομών της R.*Eco*RV και R.*Eco*RI συμπλοκοποιημένων με την συγγενή τους αλληλουχία DNA.

Οι δύο ενδονουκλεάσες των οποίων οι δομές είναι γνωστές R.*Eco*RI και R.*Eco*RV δεν εμφανίζουν σημαντική ομοιότητα σε επίπεδο πρωτεϊνικής αλληλουχίας {Bogueleret et al., 1984}, οι δομές τους όπως θα περίμενε ίσως κανείς δεν εμφανίζουν ομοιότητες (εικόνα 5, εικόνες 2α,2β). Ο τρόπος προσέγγισης του DNA στις δύο περιπτώσεις είναι σημαντικά διαφορετικός. Στην περίπτωση της

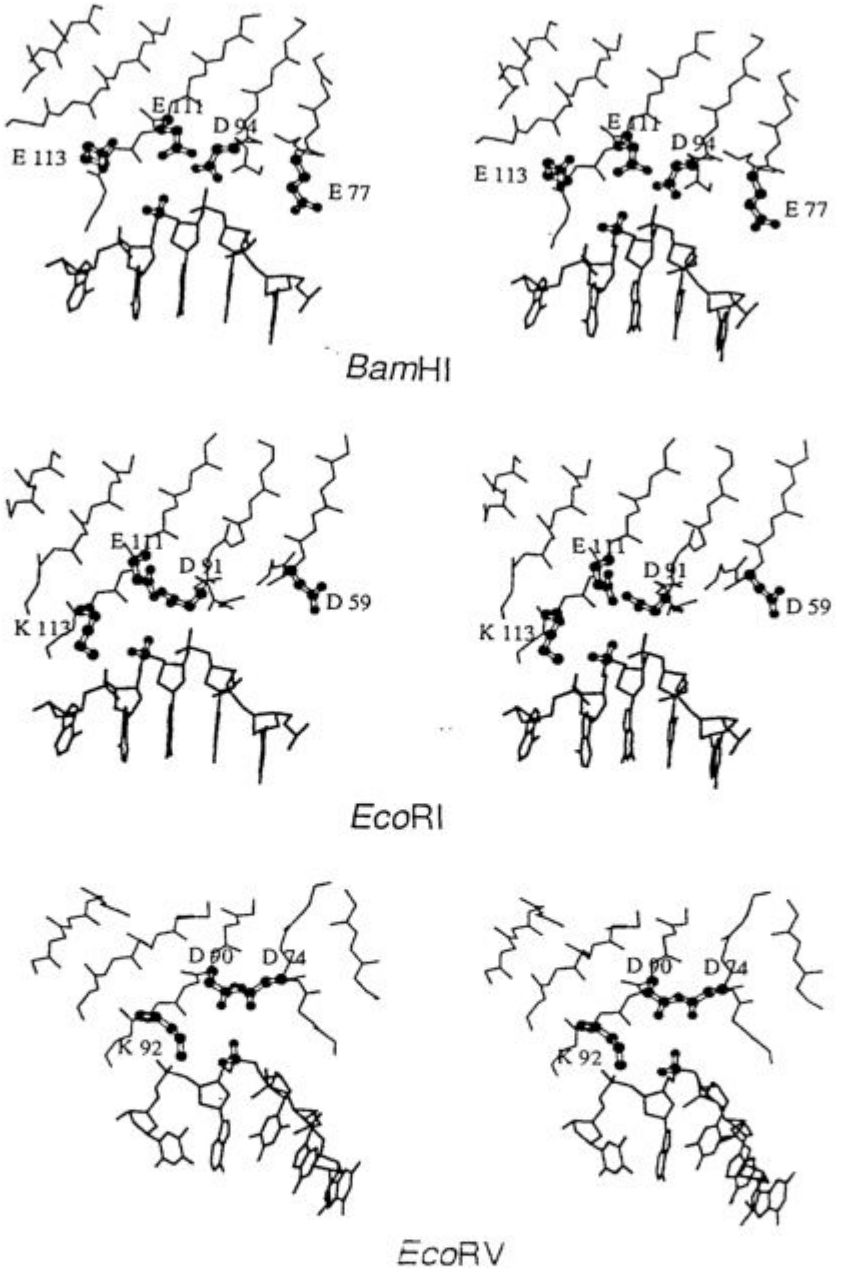
R.*Eco*RI η προσέγγιση του DNA γίνεται από την πλευρά του κύριου αυλακιού του DNA ενώ οι ειδικές επαφές γίνονται επίσης στο κύριο αυλάκι της διπλής έλικας του DNA από κατάλοιπα που ανήκουν σε ένα 4-α-ελικοειδές δεμάτι (4-α-helix bundle) {McClarin et al., 1986}, {Kim et al., 1990}. Σε αντίθεση η R.*Eco*RV πλησιάζει το DNA στην κατεύθυνση του μικρού αυλακιού και οι ειδικές επαφές πραγματοποιούνται από ευκίνητους βρόγχους της πρωτεΐνης τόσο στο μικρό όσο και στο μεγάλο αυλάκι του DNA {Winkler et al., 1993}.

Η δομή της ελεύθερης μορφής της R.*Eco*RV και η σχέση της με την δομή του συμπλόκου της πρωτεΐνης με συγγενές DNA.

Η μόνη προηγούμενα δομή ενδονουκλεάσης σε μορφή ελεύθερη από DNA ήταν αυτή της R.*Eco*RV (εικόνα 2α) {Winkler et al., 1993} και σύγκριση της μορφής αυτής με αυτήν της πρωτεΐνης δεμένης στην συγγενή αλληλουχία του DNA δείχνει ότι σημαντικές αλλαγές στην διαμόρφωση της πρωτεΐνης συμβαίνουν κατά την ειδική πρόσδεση στο DNA. Το διμερές της R.*Eco*RV είναι ένα μόριο σχήματος "U" που στην μορφή του συμπλόκου "αγκαλιάζει" το DNA ενώ οι δύο βραχίονες περιλαμβάνουν τους βρόγχους με τους οποίους γίνεται η αναγνώριση της αλληλουχίας. Το ένζυμο στην μορφή του συμπλόκου διατηρεί όλα τα δευτεροταγή στοιχεία της δομής της ελεύθερης πρωτεΐνης, μετακινώντας όμως την θέση των δύο βραχιόνων έτσι ώστε το μόριο να βρεθεί σε μία "κλειστή" δομή ικανή για την αναγνώριση και την τομή του DNA. Ακόμα και στην δομή της ελεύθερης μορφής της R.*Eco*RV το άνοιγμα πρόσδεσης του DNA παρουσιάζεται κλειστό για την είσοδο του DNA. Η ευκαμψία της τεταρτοταγούς δομής του μορίου και ειδικά των βρόγχων αναγνώρισης είναι σημαντική για το άνοιγμα της περιοχής πρόσδεσης του DNA στα αρχικά στάδια σύνδεσης της πρωτεΐνης με το DNA {Winkler et al., 1993}.

Το σύμπλοκο με μή συγγενές DNA.

Παρά το γεγονός ότι στην περίπτωση της R.*Eco*RV έχουν γίνει προσπάθειες για την κρυστάλλωση και επίλυση της δομής του συμπλόκου με μή συγγενές (τυχαίας αλληλουχίας) DNA, το σύμπλοκο το οποίο δημιουργήθηκε και του οποίου η δομή προσδιορίστηκε είναι αμφίβολο κατά πόσον μπορεί να θεωρηθεί αντιπροσωπευτικό ενός συμπλόκου με μή-συγγενές DNA καθώς, δύο οκταμερή δίκλωνα ολιγονουκλεοτίδια έχουν προσδεθεί στο διμερές της πρωτεΐνης {Winkler et al., 1993}, προσομοιάζοντας έτσι περισσότερο την δομή του συμπλόκου, μετά την τομή του DNA. Ο προσδιορισμός της δομής του συμπλόκου με το μή-συγγενές DNA ως συνεχές μόριο είναι απαραίτητη προκειμένου να βγούν συμπεράσματα για την διαμόρφωση του μορίου στο στάδιο της αναζήτησης στο μόριο DNA της αλληλουχίας αναγνώρισης με την προτεινόμενη διαδικασία της διευκολυνόμενης διάχυσης (facilitated diffusion) {Taylor et al., 1991}.



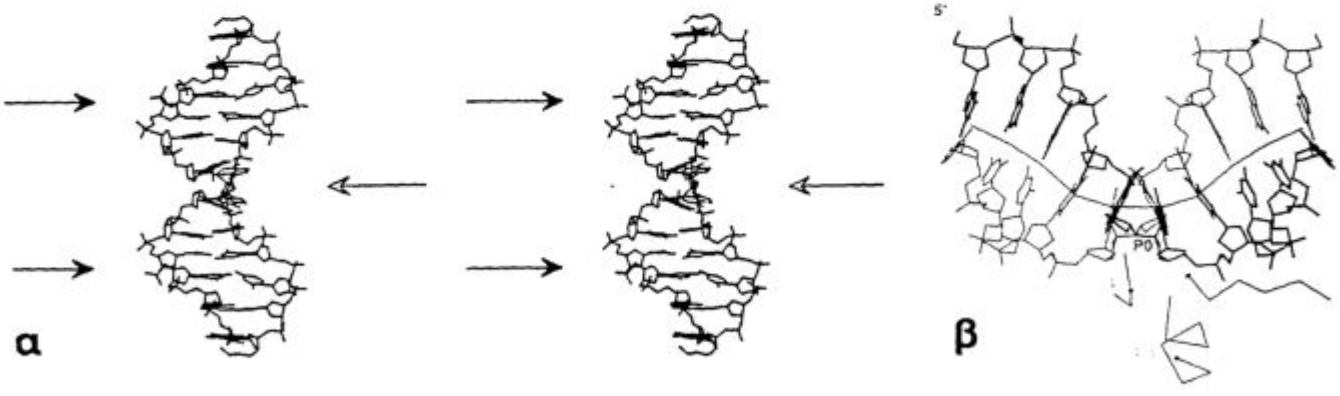
Εικόνα 3. Στερεοσκοπική παράσταση του καταλυτικού κέντρου των R.EcoRV, R.EcoRI και R.BamHI (υποθετικό μοντέλο). *Σε κάθε περίπτωση διακρίνεται στο μόριο του DNA ο διασπάσιμος φωσφοδιεστερικός δεσμός. Στον σχηματισμό του καταλυτικού κέντρου έχει υποτεθεί ότι μετέχει και το ιόν Mg^{2+} το οποίο δεν παρουσιάζεται καθώς απουσιάζει από όλες τις δομές ενδονουκλεασών μέχρι σήμερα. {Newman, et al., 1994a}*

Η καταλυτική περιοχή.

Η περιοχή των πρωτεϊνών που έχουν δειχθεί ως υπεύθυνες για την κατάλυση αντιπροσωπεύει το μόνο κοινό δομικό στοιχείο ανάμεσα στις δομές των R.*Eco*RV και R.*Eco*RI (εικόνα 3). Δύο όξινες και μία βασική πλευρική αλυσίδα αμινοξέων βρίσκονται σε κοντινή απόσταση από τον φωσφοδιεστερικό δεσμό που διασπάται {Anderson, 1993}. Στα δύο ένζυμα η καταλυτική περιοχή μπορεί να περιγραφεί με το πρότυπο **Pro,Asp,X15-19,Glu-Asp,X,Lys** (X οποιοδήποτε αμινοξικό κατάλοιπο) {Anderson, 1993}. Η συσχέτιση του προηγούμενου προτύπου με τις καταλυτικές περιοχές των περιοριστικών ενζύμων γενικότερα, είναι ασθενής, καθώς εμφανίζεται μόνο σε 11 από τις 36 αλληλουχίες περιοριστικών της SWISS-PROT βάσης δεδομένων ενώ στην αλληλουχία της R.*Eco*RI εμφανίζεται δεύτερη φορά χωρίς συσχέτιση με την καταλυτική περιοχή {Anderson, 1993}. Μελέτη και άλλων συμπλόκων ενδονουκλεασών με DNA είναι απαραίτητη για να αποδοθεί η ακριβής σημασία του προαναφερθέντος προτύπου.

Ο μηχανισμός της κατάλυσης.

Για τον μηχανισμό της κατάλυσης και τον ρόλο των ιόντων Mg^{2+} έχουν προταθεί δύο παρόμοιοι μηχανισμοί βασισμένοι σε μηχανισμούς άλλων μη ειδικών νουκλεασών με απαίτηση δισθενών ιόντων {Selent et al., 1992}, {Jeltsch et al., 1992}. Και στους δύο μηχανισμούς προτείνεται μια νουκλεοφιλική υποκατάσταση τύπου S_N2 από ενεργοποιημένο μόριο νερού στην φωσφορική ομάδα. Στην μία περίπτωση



Εικόνα 4. Παράσταση της στρέβλωσης του DNA α) στο σύμπλοκο της EcoRI {McClarin et al., 1986} και β) EcoRV {Winkler et al., 1993}. *α) Το ανοικτό βέλος δείχνει στο κέντρο της τύπου I νεοκάμψης ενώ τα μαύρα βέλη στίς τύπου II νεοκάμψεις που λόγω συμμετρίας είναι πανομοιότυπες. β) PO ο διασπάσιμος δεσμός. Η γραμμή που ακολουθεί το μονοπάτι της έλικας σημειώνεται για να φανεί η κάμψη του μορίου.*

η ενεργοποίηση του μορίου νερού επιτυγχάνεται από το Mg^{2+} και μία από τις όξινες πλευρικές αλυσίδες (Asp,Glu) ενώ στον δεύτερο μηχανισμό στην ενεργοποίηση μετέχει η γειτονική φωσφορική ομάδα της ραχοκοκκαλιάς του DNA. Στο ενδιάμεσο στάδιο δημιουργείται μία πεντασθενής διπλά αρνητική φωσφορική ομάδα η οποία σταθεροποιείται από το βασικό αμινοξύ (Lys) και το ιόν του Mg^{2+}.

Η δομή του δεσμευμένου σε ενδονουκλεάσες DNA.

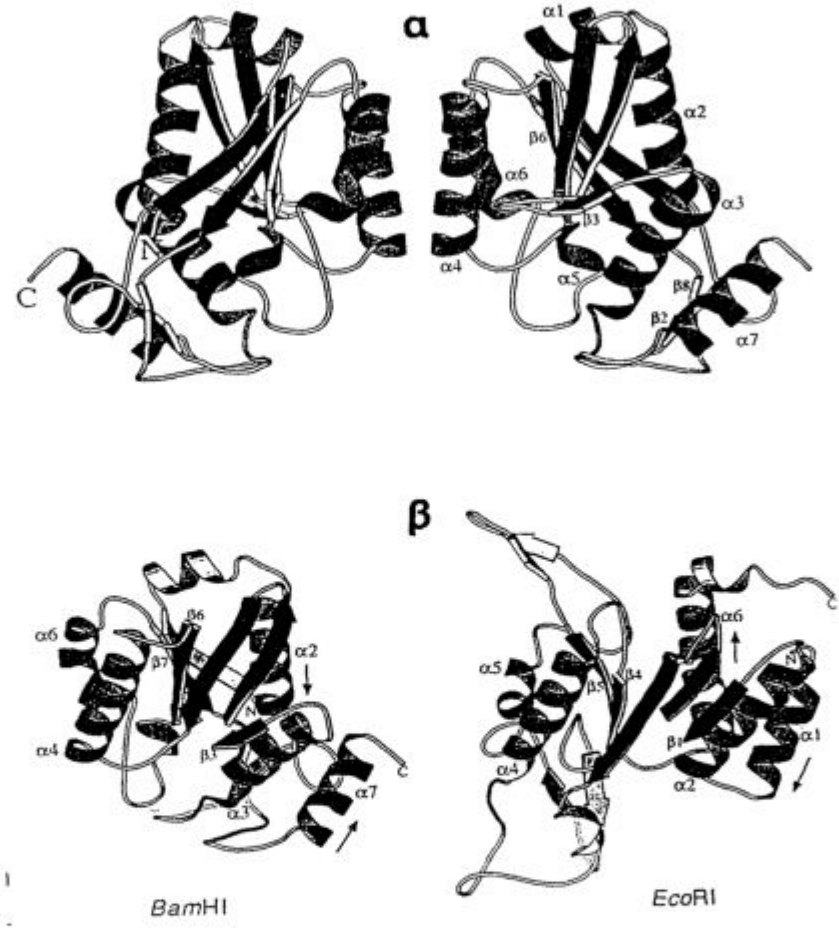
Σημαντική απόκλιση από την δομή του DNA σε B-μορφή παρουσιάζει το DNA τόσο προσδεδεμένο στην R.*Eco*RI όσο και στην R.*Eco*RV (εικόνες 4α,4β). Στην περίπτωση της R.*Eco*RV μία κεντρική (σε σχέση με τον τάξεως-2 άξονα συμμετρίας) κάμψη 50° είναι το πιό ενδιαφέρον χαρακτηριστικό {Winkler et al., 1993}. Η αλλοίωση της δομής κάνει το κύριο αυλάκι της έλικας του DNA στενότερο και βαθύτερο σε σχέση με την δομή τυπικού B-DNA και προσεγγίζει την A-μορφή DNA. Η διαφοροποίηση αυτή αφορά μόνο την κεντρική περιοχή του μορίου ενώ το υπόλοιπο μόριο ταιριάζει ικανοποιητικά με το μοντέλο της B-μορφής DNA. Διαφορετική είναι η αλλοίωση της δομής του DNA στο σύμπλοκο της R.*Eco*RI, δύο τύποι κάμψης χαρακτηρισμένοι ως τύπου I και II νεοκάμψεις προκαλούν σημαντική στρέβλωση της έλικας του DNA {Frederick et al. 1985}, {McClarin et al. 1986}. Η τύπου I νεοκάμψη οδηγεί σε ένα ξεδίπλωμα της διπλής έλικας του DNA 25° με επίκεντρο το σημείο διέλευσης του τάξης-2 άξονα συμμετρίας του μορίου και μπορεί να έχει επίδραση στο DNA σε απόσταση από το σημείο πρόσδεσης της πρωτεΐνης.

Στην περίπτωση και των δύο πρωτεϊνών έχει υποτεθεί ότι η ενέργεια που απαιτείται για την στρέβλωση του DNA προέρχεται από την δημιουργία ειδικών και μή αλληλεπιδράσεων πρωτεΐνης-DNA. Η συμμετοχή όπως αναφέρθηκε φωσφορικών ομάδων της ραχοκοκκαλιάς του DNA στην δημιουργία της περιοχής πρόσδεσης των ιόντων Mg^{2+} και του καταλυτικού κέντρου καθιστά την στρέβλωση του DNA αναγκαία συνθήκη για την καταλυτική λειτουργία στην περίπτωση των δύο αυτών ενζύμων {Winkler et al., 1993}.

Η δομή της ελεύθερης μορφής της *Bam*HI ενδονουκλεάσης

Ταυτόχρονα με την επίλυση της δομής της R.*Pvu*II πραγματοποιήθηκε ο προσδιορισμός της δομής της ελεύθερης μορφής της *Bam*HI (εικόνα 5α) ενδονουκλεάσης σε 1.95 Å διακριτικότητα με την μέθοδο ανώμαλης σκέδασης με πολλαπλά μήκη κύματος (Multiple Anomalous Diffraction-MAD) {Newman et al., 1994a}. Κάθε υπομονάδα της R.*Bam*HI δείχθηκε να αποτελείται από μία κεντρική β-πτυχωτή επιφάνεια περιβαλλόμενη από α-έλικες. Η δομή του διμερούς σχηματίζει ένα μεγάλο αυλάκι το οποίο θα μπορούσε να δεχθεί ενα μόριο B-DNA.

Σημαντική είναι η ομοιότητα που παρατηρείται ανάμεσα στην δομή της R.*Bam*HI και αυτής της R.*Eco*RI {Newman et al., 1994b} παρά την απουσία οποιασδήποτε σημαντικής ομοιότητας σε επίπεδο αμινοξικής αλληλουχίας (εικόνα 5β). Ενα κοινό κεντρικό δομικό πρότυπο (common core motif: CCM) ορίζεται με βάση την ομοιότητα των δύο ενζύμων που τέμνουν το DNA αφήνοντας 5' προεξέχοντα άκρα.



Εικόνα 5. Παράσταση κορδέλας της δομής του διμερούς της R.BamHI (α). Τοπολογικές ομοιότητες μεταξύ των R.BamHI και R.EcoRI (β). (Newman et al., 1994)

Α 4. Το *Pvu*II RM σύστημα.

Το 1981 ανακαλύφθηκαν {Gingeras et al., 1981} στο βακτήριο *Proteus vulgaris* δύο συστήματα περιορισμού και τροποποίησης του DNA. Το σύστημα με ειδικότητα για την αλληλουχία $^{5'}$CAG/CTG$_{3'}$ ονομάστηκε *Pvu*II ενώ το δεύτερο RMs με αλληλουχία αναγνώρισης $^{5'}$CGAT/CG$_{3'}$, ονομαστήκε *Pvu*I και κλωνοποιήθηκε αργότερα {Smith et al., 1992}. Για το περιοριστικό ένζυμο (R.*Pvu*II) του *Pvu*II συστήματος δείχθηκε ότι είναι μια πρωτεΐνη μεγέθους 18kDa περίπου η οποία τέμνει το DNA στο μέσο της αλληλουχίας αναγνώρισης αφήνοντας τμήματα DNA με "τυφλά" άκρα (blunt end). Το 1985 από τους Blumenthal et al. κλωνοποιήθηκε μια περιοχή μεγέθους ≈ 3.4kb, από ένα πλασμίδιο με την ονομασία pPVU1, υπεύθυνη για την κωδικοποίηση των πρωτεϊνών του *Pvu*II συστήματος. Χαρτογράφηση της περιοχής αυτής με την δημιουργία ελλείψεων προσδιόρισε τις θέσεις των γονιδίων των R.*Pvu*II και M.*Pvu*II. Τα σημεία έναρξης των κωδικών περιοχών των δύο γονιδίων δείχθηκε ότι απέχουν 200 bp περίπου και η μεταγραφή τους πραγματοποιείται σε αντίθετες κατευθύνσεις. Η διαγονιδιακή περιοχή των 200bp περιέχει όλα τα cis στοιχεία απαραίτητα για την μεταγραφή τους. Το 1991 από το ίδιο εργαστήριο δείχθηκε η παρουσία ενός ανοικτού πλαισίου διαβάσματος στην διαγονιδιακή περιοχή το οποίο κωδικοποιεί μία πρωτεΐνη (C.*Pvu*II) μεγέθους 84 καταλοίπων, κατά την παρουσία της οποίας η μεταγραφή του γονιδίου της R.*Pvu*II αυξάνεται κατά 10^5 φορές {Tao & Blumenthal, 1992}. Ρυθμιστικές πρωτεΐνες ομόλογες με την C.*Pvu*II ανακαλύφθηκαν και σε άλλα RMs τύπου II, παρά την απουσία ομολογιών μεταξύ των περιοριστικών ενζύμων των συστημάτων αυτών {Tao, Bourne and Blumenthal, 1991}.

Α 5. Σκοπός της περιγραφόμενης εργασίας.

Σκοπός της εργασίας υπήρξε ο κρυσταλλογραφικός προσδιορισμός της δομής της *Pvu*II ενδονουκλεάσης στην μορφή ελεύθερης πρωτεΐνης. Η επιλογή της συγκεκριμένης ενδονουκλεάσης έγινε πέρα από πρακτικούς λόγους (διαθεσιμότητα κλώνου) με βάση το γεγονός ότι το *Pvu*II σύστημα είναι από τα πλέον μελετημένα σε βιοχημικό και μοριακό επίπεδο πράγμα που επιτρέπει, συνδυάζοντας πολύπλευρη γνώση, την βαθύτερη κατανόηση των ιδιοτήτων του. Παράλληλα η *Pvu*II ενδονουκλεάση όντας το μικρότερο σε μέγεθος γνωστό ένζυμο της κατηγορίας του {Wilson & Murray, 1991} περιέχει την απόλυτα απαραίτητη πληροφορία που απαιτείται για την αναγνώριση του DNA και την καταλυτική δράση ενός

περιοριστικού ενζύμου. Τέλος η R.*Pvu*II όπως και η R.*Eco*RV κόβουν την αλυσίδα του DNA κατά τον ίδιο τρόπο αφήνοντας τμήματα DNA με "τυφλά" άκρα. Με την εργασία αυτή δίνεται η δυνατότητα για τον έλεγχο της υπόθεσης ότι τα περιοριστικά ένζυμα μπορούν να ομαδοποιηθούν δομικά με βάση την θέση του διασπάσιμου δεσμού {Anderson, 1993}. Η επιλογή της ελεύθερης πρωτεΐνης (αντί του συμπλόκου με DNA) για την επίλυση της δομής της R.*Pvu*II έγινε γιατί οι κρύσταλλοι της ελεύθερης πρωτεΐνης ήταν οι πρώτοι (της R.*Pvu*II) που αποκτήθηκαν στο εργαστήριο που πραγματοποιήθηκε η εργασία και με την προοπτική η πληροφορία από την δομή της ελεύθερης πρωτεΐνης ενδεχόμενα να χρησιμοποιηθεί μελλοντικά για τον προσδιορισμό της δομής του συμπλόκου της πρωτεΐνης με DNA, με την μέθοδο της μοριακής αντικατάστασης (molecular replacement). Στην πορεία της εργασίας αυτής έγινε γνωστή η εργασία άλλης ερευνητικής ομάδας πάνω στον προσδιορισμό της δομής του συμπλόκου της πρωτεΐνης με συγγενές DNA {Balendiran at al., 1994}, {Cheng et al., 1994} η οποία ολοκληρώθηκε περίπου ταυτόχρονα με αυτήν της ελεύθερης πρωτεΐνης. Σύγκριση των δύο δομών που αποκαλύπτει τις αλλαγές στη διαμόρφωση του μορίου κατά την πρόσδεση στο DNA γίνεται στο σχετικό κεφάλαιο των αποτελεσμάτων.

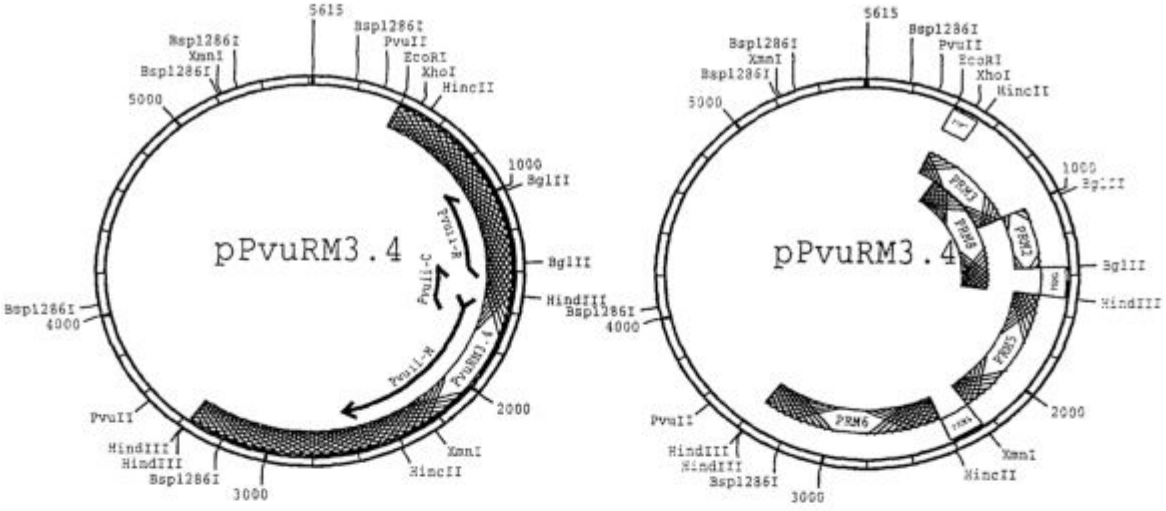
Β. Αποτελέσματα και Συζήτηση

Β 1. Προσδιορισμός και ανάλυση της νουκλεοτιδικής αλληλουχίας του *Pvu*II συστήματος.

Προκειμένου να προχωρήσουμε σε μια κρυσταλλογραφική δομική ανάλυση είναι αναγκαία η γνώση της πρωτοταγούς δομής της πρωτεΐνης, γεγονός που μας οδήγησε στο να πραγματοποιήσουμε τον προσδιορισμό της αλληλουχίας της περιοχής του DNA που κωδικοποιεί το *Pvu*II RMs. Συνολικά προσδιορίστηκε η αλληλουχία ενός τμήματος 2.7kb από τις 3.4kb του κλώνου p*Pvu*RM3.4 που περιέχει το σύνολο της πληροφορίας για το *Pvu*II σύστημα {Blumenthal et al., 1985} και έχει αναφερθεί στο αντίστοιχο τμήμα της εισαγωγής. Στην περιοχή αυτή εκτός από το περιοριστικό ένζυμο, κωδικοποιούνται η *Pvu*II μεθυλάση καθώς και η C.*Pvu*II πρωτεΐνη που έχει ρυθμιστικό ρόλο στην έκφραση του περιοριστικού ενζύμου {Tao & Blumenthal, 1992}. Τέλος στην μεγέθους 200 βάσεων περιοχή μεταξύ των R.*Pvu*II και M.*Pvu*II περιλαμβάνονται τα απαραίτητα cis στοιχεία για την μεταγραφή του συστήματος {Blumenthal et al., 1985}, {Athanasiadis et al., 1990}. Πέρα από την αναγκαιότητα της γνώσης της αλληλουχίας της πρωτεΐνης για τον προσδιορισμό της δομής, οι συγκρίσεις των αλληλουχιών με γνωστές αλληλουχίες πρωτεϊνών σε βάσεις δεδομένων, τόσο του περιοριστικού ενζύμου όσο και της μεθυλάσης του συστήματος, αποτελούν σημαντικά στοιχεία για την κατανόηση της εξελικτικής σχέσης των δύο πρωτεϊνών τόσο μεταξύ τους όσο και με πρωτεΐνες άλλων συστημάτων περιορισμού και τροποποίησης του DNA. Τέλος η γνώση της αλληλουχίας του υπο μελέτη ενζύμου μας έδωσε την δυνατότητα εκτιμήσεων για τον ειδικό σχεδιασμό και την επιλογή των ενώσεων βαρειών ατόμων που χρησιμοποιήθηκαν για την μέθοδο της ισόμορφης αντικατάστασης σε μετέπειτα στάδια της εργασίας.

Β 1.1 Στρατηγική του προσδιορισμού της νουκλεοτιδικής αλληλουχίας.

Για τον προσδιορισμό της αλληλουχίας χρησιμοποιήθηκε ο κλώνος pPvuRM3.4 και μια σειρά υποκλώνων του, από ανεξάρτητη κλωνοποίηση που πραγματοποιήθηκε στο εργαστήριο Μοριακής Βιολογίας Θηλαστικών του Καθ. Ι. Παπαματθαιάκη από τους Δρ. Δ. Θάνο και Δρ. Μ. Γρηγορίου {Αδημοσίευτα αποτελέσματα}. Στην εικόνα 6 περιγράφεται ο χάρτης του ενθέματος και του πλασμιδίου φορέα του κλώνου καθώς και οι θέσεις περιορισμού που χρησιμοποιήθηκαν για την κατασκευή, των υποκλώνων με την ονομασία Prm1-Prm8. Τόσο ο συνολικός κλώνος όσο και οι υποκλώνοι έχουν εντεθεί σε φορείς pUC19/18. Ο προσδιορισμός της αλληλουχίας έγινε σε δύο στάδια, πρώτα



Εικόνα 6. Χάρτης του pPvuRM3.4 πλασμιδίου. *Το πλασμίδιο pUC19 με το PvuII RMs κλωνοποιημένο στις EcoRI-HindIII θέσεις. α) Οι περιοχές των κωδικοποίησης των τριών πρωτεϊνών του PvuII συστήματος (βέλη). β) Οι υποκλώνοι Prm1-Prm8 (σκιασμένες περιοχές).*

για κάθε υποκλώνο πραγματοποιήθηκε προσδιορισμός της αλληλουχίας αμφίδρομα με την χρήση των καταλλήλων μορίων εκκινητών (dideoxy-sequencing), ειδικών για ενθέματα στα pUC19/18 πλασμίδια, και περιγράφεται στο κεφάλαιο των μεθόδων. Στην δεύτερη φάση η αλληλουχία η οποία προσδιορίστηκε με τον προηγούμενο τρόπο χρησιμοποιήθηκε για τον σχεδιασμό ολιγονουκλεοτιδίων εκκινητών (πίνακας 2, εικόνα 7). Αυτό έγινε για την επαλήθευση της αλληλουχίας εκεί όπου το μέγεθος των υποκλώνων δεν επέτρεπε την επικάλυψη των αλληλουχιών που διαβάστηκαν με την χρήση των εκκινητών του pUC19/18, καθώς και για τις περιοχές στα άκρα των υποκλώνων για την αποφυγή σφαλμάτων λόγω κοντινών θέσεων περιορισμού κατά την διαδικασία της υποκλωνοποίησης. Η ανάγνωση της αλληλουχίας με τα ολιγονουκλεοτίδια O1-7 έγινε με υπόστρωμα το πλασμίδιο pPvuRM3.4.

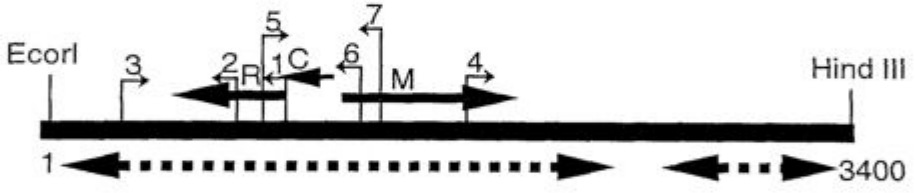
Β 1.2. Ανάλυση της νουκλεοτιδικής αλληλουχίας της R.*Pvu*II

Εχοντας προσδιορίσει την αλληλουχία πού φαίνεται στον πίνακα 4 μπορέσαμε να εντοπίσουμε τρία ανοιχτά πλαίσια διαβάσματος, τα δύο από τα οποία ήταν σύμφωνα με τα δεδομένα από ανάλυση ελλείψεων για την θέση των R.*Pvu*II και M.*Pvu*II γονιδίων {Blumenthal et al., 1985}. Το τρίτο ανοιχτό πλαίσιο

Πίνακας 2. Κατάλογος ολιγονουκλεοτιδίων εκκινητών που χρησιμοποιήθηκαν στην δεύτερη φάση προσδιορισμού της αλληλουχίας του *PvuII* RM συστήματος.

Ονομα	Αλληλουχία	Θέση και φορά	Περιοχή Ανάγνωσης
01	TAGCAAATGCCTTAAATG	1046-1029	PRM1->PRM2*
02	AGAAGGTAATGATGCTGT	835-818	μέσο PRM2
03	CTTTCATCGGTTTAGGCG	294-311	μέσο PRM3
04	AAGACCCAGACAAATTCT	1810-1828	PRM5->PRM4->PRM6*
05	TAAGTCTTGATATTCCTG	930-948	PRM2->PRM1*
06	ACCCATTACTAGTTGTGT	1330-1312	M-Αρχή§
07	CATTACGCTTCCCTCCTG	1420-1402	M-Αρχή§

* Με βέλη συμβολίζονται τα σημεία ένωσης των υποκλώνων.
§ M: Μεθυλάση



Εικόνα 7. Θέση και διεύθυνση ανάγνωσης των ολιγονουκλεοτιδίων Ο1-Ο7. *Με διακεκομμένα βέλη δείχνεται η περιοχή του pPvuRM3.4 της οποίας προσδιορίστηκε η αλληλουχία ενώ με συνεχόμενα τα τρία γονίδια του συστήματος. (R : R.PvuII, M : M.PvuII, C . C PvuII)*

διαβάσματος C.*Pvu*II, αργότερα αναγνωρίστηκε {Tao & Blumenthal, 1992}, {Tao et al., 1991} ώς παράγοντας ρύθμισης της μεταγραφής του R.*Pvu*II. Οι τρείς πρωτεϊνικές αλληλουχίες μεγέθους 157, 336 και 84 αμινοξέων αντίστοιχα παρουσιάζονται επίσης, στον πίνακα 4. Η περιοχή της αλληλουχίας 1-1321 {Athanasiadis et al., 1990} κατατέθηκε στην EMBL βάση δεδομένων με τον κωδικό pvpvuii και αριθμό πρόσβασης X52681. Το υπόλοιπο τμήμα της αλληλουχίας που ταυτοποιήθηκε και περιλαμβάνει την αλληλουχία της *Pvu*II μεθυλάσης προσδιορίστηκε ανεξάρτητα από άλλο εργαστήριο {Tao et al., 1989} και μπορεί να

αναζητηθεί στην βάση δεδομένων EMBL με τον κωδικό Prmpvuiib. Διαφοροποίηση μίας βάσης έχει ως αποτέλεσμα να εμφανίζεται μια συντηρητική αλλαγή μεταξύ των δύο ανεξάρτητα προσδιορισμένων αλληλουχιών της μεθυλάσης : 45D=>E.

Αναζήτηση ομολογιών μεταξύ της R.*Pvu*II και των γνωστών μέχρι σήμερα πρωτεϊνικών αλληλουχιών σε βάσεις δεδομένων (Swiss-Prot, PIR-Protein) δείχνει ομολογίες κάτω από τα όρια στατιστικής σημαντικότητας. Ενδεικτικά η στοίχιση της καλύτερης αρχικής ομοιότητας από ανίχνευση της Swiss-Prot (κωδικός εγγραφής sw:P05548, πρόγραμμα Fasta {Devereux, Haeberli and Smithies, 1984}) παρουσιάζει 18.24% ταυτότητα αμινοξικών καταλοίπων (πρόγραμμα Bestfit {Devereux, Haeberli and Smithies, 1984}), με δείκτη ποιότητας της στοίχισης Q=50.1 και με μέσο δείκτη ποιότητας, ύστερα από τυχαιοποίηση της αλληλουχίας 10 φορές, Qr=51.7 +/- 2. Το μοριακό βάρος της πρωτεΐνης (18345 Da), όπως υπολογίζεται από την αλληλουχία, είναι σε πλήρη συμφωνία με το φαινόμενο μοριακό βάρος σε αποδιατακτικό πήκτωμα ακρυλαμίδης (κεφάλαιο B3, εικόνα 10). Προσδιορισμός της αλληλουχίας του αμινοτελικού άκρου της R.*Pvu*II πρωτεΐνης {Tao & Blumenthal, 1992} δείχνει την απομάκρυνση της Met1 in vivo ενώ είναι σε πλήρη συμφωνία με τα επόμενα 11 κατάλοιπα. Στην ίδια εργασία παρουσιάζεται ανεξάρτητος προσδιορισμός της νουκλεοτιδικής αλληλουχίας του γονιδίου της R.*Pvu*II και είναι, για την κωδική περιοχή, σε πλήρη συμφωνία με αυτόν του πίνακα 4. Το θεωρητικό ισοηλεκτρικό σημείο της πρωτεΐνης υπολογίστηκε με το πρόγραμμα Isoelectric {Devereux, Haeberli and Smithies, 1984} και βρέθηκε να είναι σε pH 6.73. Στον πίνακα 3 παρουσιάζεται ανάλυση της αμινοξικής σύνθεσης της πρωτεΐνης. Χαρακτηριστικό της αλληλουχίας της R.*Pvu*II η εμφάνιση διπλάσιας περιεκτικότητας καταλοίπων His (5.1%) από τη μέση σύσταση των πρωτεϊνών ενώ τα αμινοξικά κατάλοιπα που φέρουν αρωματικούς δακτυλίους, στο σύνολο τους, εμφανίζονται επίσης σε ψηλό ποσοστό 16.5% (μέση σύσταση πρωτεϊνων 10.8% {Nakashima, Nishikawa and Ooi, 1986})

Πίνακας 3. Αμινοξική σύσταση της R.*Pvu*II.

A	7	D	11	E	10	F	4
G	11	H	8	I	16	K	13
L	16	M	3	N	9	P	8
Q	6	R	5	S	4	T	5
V	7	W	4	Y	10		

Πίνακας 4. Νουκλεοτιδική αλληλουχία του PRM3.4 ενθέματος. Υπογραμμισμένη η πρωτεινική αλληλουχία της R *Pvu*II και με πλάγια γράμματα της C.*Pvu*II πρωτείνης. Με αστερίσκο σημειώνονται τα σημεία λήξης. Οι χαρακτήρες b,d και f υποδηλώνουν το παράθυρο διαβάσματος. Με περίγραμμα σημειώνονται τα εναλλακτικά σημεία εκκίνησης της μετάφρασης για την μεθυλάση {Tao et al., 1989}. Τέλος με έντονο χαρακτήρα είναι το αμινοξύ που αποτελεί την μοναδική συντηρητική αλλαγή (45D->E) που παρατηρήθηκε μεταξύ των δύο ανεξάρτητα προσδιορισμένων αλληλουχιών της M.*Pvu*II.

```
    1   TGACCGAGACAGAACAACTCTTAGAGAACGCCTTGAACACCTTGTTACAACAGCAAGAGA
   61   CCGTTATGCAGCAATTTATCTCAGCGTGTCGGGGGAACTCGAGACACAGAACAGTCAGTT
  121   AGCCATGCAATACAGGGAAATCAGTCAACAGTAGCGACTTTAACCGGACTGCTGAACAGG
  181   TATCAGAACGGCTGACACAGCAGATCCAGCGACAGGAGCAGAGTCGTAGTCGGGGGTTTC
  241   AGTATCGGTTTTTAACAGGATAAAACCGCTTCAAGCCACAAACAAGTTTGCGACTTTCAT
  301   CGGTTTAGGCGATCAAAGTCGGACGAAAAACGGATGTAAACAGCCCTCAAACGTCACAGA
  361   AACCGTTATAAACTTTCAGGATGGATTTGGGTGTTATGGGGGGTGTTTGGTGCTTAAAAT
  421   CGTGCTGATGCGTTCTGATGGTGACTGACGGGCGGTTTTAAGGTGTAAACTCGCCCGTTT
  481   AACATAATTGGGTCTTATCGGAACCATGAATGTAGCTCCAATTTAGTAAATCTTTGTCCC
                                                          * Y I K T G
  541   ATGTTCCATTACATATTTTACAGGTATTTTAGGGTTGTTAATATCTTTATGCCCATCTGA
f        H E M V Y K V P I K P N N I D K H G D S
  601   ATACCATTTCCTTTCCCATTTATCATAGTAAAATTCTAGATCTTTTGGCTCTAATCTGTA
f        Y W K A E W K D Y Y F E L D K P E L A Y
  661   TATAGCTTCTATTGCGATACCACGGTATATGGCAAAAATCCAAGGTACTTGTCTATATTT
f        I A E I A I G A Y I A F I W P V Q A Y K
  721   TGCAATAATTACAGGATTCATGTGGTGGTGAGTTGAAAAACCTTTAGTGAGGTCTATGTT
f        A I I V P N M H H H T S F G K T L D I N
  781   TATTGATTTTAACTCGTATTCTTGTCCTGCGTTATCTACAGCATCATTACCTTCTCGTCC
f        I S K L E Y E Q G A N D V A D N G E R G
  841   TGGTAGTACTGTTAATCCTGTAATTAGAAGGACTTGAAGCAACTTTCCACCATTATCTTG
f        P L V T L G T I L L V Q L L K G G N D Q
  901   AAAAATATCATTTATTCCATGTTTTAATGCTAAGTCTTGATATTCCTGTATATGCGGCCA
f        F I D N I G H K L A L D Q Y E Q I H P W
  961   AAGCTCTAATAATTTATTTAGATCTGGGTGACTCATTTTCGTGTTCCATCATTAGTATTG
d                     * K S A P S E N E H E M M L I
f        L E L L K N L D P H S M
                                        <-- R.PvuII
 1021   ATATAGAAACATTTAAGGCATTTGCTATTCGCTCAATGTTGTCTATCGATATATTCCTTT
d       S I S V N L A N A I A E I N D I S I N A
 1081   CCGCTCGTTCAATTGAACCAATGTAGGTTCTATGGATTCCCACTAGATCAGCTAAGCTTT
d       E A A E I S G I Y T A H I G V L D A L S
 1141   CTTGGGATAGACCTAGCTCGCCTCGCATTTTTTTTACATTCTTTGCAAGGGTTAATCGGG
d       E Q S L G L E G A M K K V N K A L T L A
 1201   CTGATAAAGGATTTGTATTTCTGCTCATAGACGGGATAATGATGGTATGATGACTTTGAA
b                                                   M M T L N
                                            M.PvuII-->
d       A S L P N T N A S M
                                    <-- C.PvuII
 1261   TCTACAGACTATGAGTAGCAATGATATGTTGAATTTTGGAAAAAAACCTGCCTACACAAC
b        L Q T M S S N D M L N F G K K P A Y T T
 1321   TAGTAATGGGTCTATGTATATAGGTGACTCATTGGAGCTATTAGAATCATTCCCAGATGA
b        S N G S M Y I G D S L E L L E S F P D E
 1381   AAGTATTAGTCTGGTTATGACTAGTCCTCCCTTCGCATTACAACGAAAAAAAGAATACGG
b        S I S L V M T S P P F A L Q R K K E Y G
 1441   AAACTTAGAACAGCATGAGTATGTGGATTGGTTTTTGTCATTTGCTAAAGTGGTTAATAA
b        N L E Q H E Y V D W F L S F A K V V N K
 1501   GAAACTAAAACCGGATGGAAGTTTTGTTGTAGATTTTGGTGGGGCGTACATGAAAGGCGT
```

```
b          K  L  K  P  D  G  S  F  V  V  D  F  G  G  A  Y  M  K  G  V
1561   TCCAGCAAGAAGTATCTATAACTTTAGAGTTTTGATCAGAATGATAGACGAAGTGGGTTT
b          P  A  R  S  I  Y  N  F  R  V  L  I  R  M  I  D  E  V  G  F
1621   TTTTCTTGCTGAAGATTTTTATTGGTTTAACCCATCAAAATTACCAAGCCCTATAGAATG
b          F  L  A  E  D  F  Y  W  F  N  P  S  K  L  P  S  P  I  E  W
1681   GGTTAATAAGAGAAAAATAAGAGTTAAAGATGCAGTGAATACAGTATGGTGGTTCTCTAA
b          V  N  K  R  K  I  R  V  K  D  A  V  N  T  V  W  W  F  S  K
1741   AACAGAGTGGCCTAAGTCTGATATAACGAAAGTTCTAGCTCCATATAGCGATAGAATGAA
b          T  E  W  P  K  S  D  I  T  K  V  L  A  P  Y  S  D  R  M  K
1801   AAAGTTAATTGAAGACCCAGACAAATTCTATACACCAAAAACTAGACCATCTGGGCATGA
b          K  L  I  E  D  P  D  K  F  Y  T  P  K  T  R  P  S  G  H  D
1861   TATAGGGAAGTCGTTCTCTAAAGACAATGGAGGTTCAATTCCTCCTAATTTATTACAAAT
b          I  G  K  S  F  S  K  D  N  G  G  S  I  P  P  N  L  L  Q  I
1921   ATCAAACTCAGAATCAAATGGTCAATATTTAGCCAATTGTAAGTTGATGGGTATTAAAGC
b          S  N  S  E  S  N  G  Q  Y  L  A  N  C  K  L  M  G  I  K  A
1981   GCATCCCGCTAGATTCCCTGCTAAGTTACCCGAGTTTTTTATTAGAATGTTAACTGAGCC
b          H  P  A  R  F  P  A  K  L  P  E  F  F  I  R  M  L  T  E  P
2041   AGATGATCTTGTTGTTGATATTTTTGGCGGCTCTAATACTACTGGTCTAGTCGCAGAAAG
b          D  D  L  V  V  D  I  F  G  G  S  N  T  T  G  L  V  A  E  R
2101   AGAATCTCGTAAATGGATCTCGTTTGAGATGAAACCTGAATATGTTGCGGCATCGGCGTT
b          E  S  R  K  W  I  S  F  E  M  K  P  E  Y  V  A  A  S  A  F
2161   TCGTTTTTTAGACAATAATATAAGTGAAGAAAAAATAACTGATATTTATAATCGAATCCT
b          R  F  L  D  N  N  I  S  E  E  K  I  T  D  I  Y  N  R  I  L
2221   TAATGGTGAGTCTTTGGATTTAAACTCTATTATCTAAATTTTAACTCGACCTTTTTGATA
b          N  G  E  S  L  D  L  N  S  I  I  *
2281   AAAAAGTAGTTCCCTGTAGGGCTTAGATCAGGTAGTCTTAGATATGTAAAAAAAAAACCG
2341   CCTCAGCAAAGGCGTTTTCTTAAAGATTGGTTCGT.CCGGTACAGGAAGCTGAGAGATAG
2401   CGCCCCAAAGACAGAAAGAGCGGATTAAAACCGGCTGTAAACTTGTAAGGTGGGAACAGG
2461   AGAGCGCACGAGGGAGCCGTCAGAGGGTTTCTTCTGGCATCTTTATAGTCCTGTCAGGTT
2521   TCGCCTATCTTGACTAGAGCACCGAAAATCAGTCGATTTTTGTGAAGCTCGTCAGGAGGC
2581   GGAGCCTATTGTAAAAGGGCTACTCTCCTCCTAGCTTTGCGCTCAGTTGTTGGTGTAGTG
2641   ACTCTGCATGCTCGTGGAGTTCTTCGCAGAGGTTGGTTTCTGCATCGAGTTTATCTGAAT

2701   TAAGTTCATTAAG
```

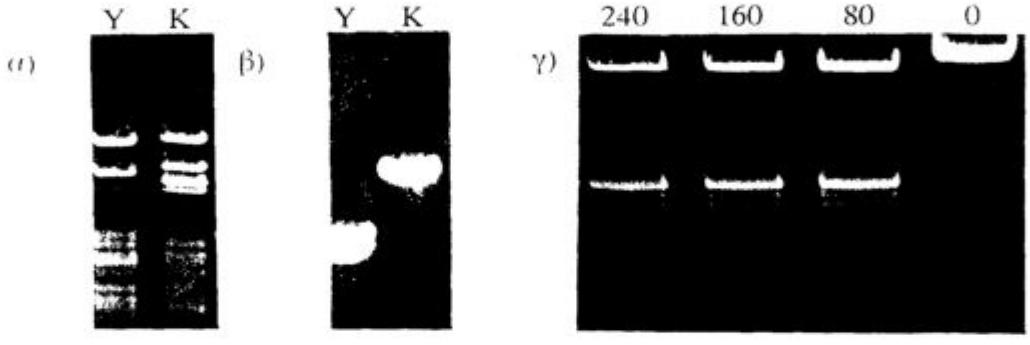
Β 2. Πρόβλεψη της αλληλουχίας της "star" (*) ενεργότητας της *Pvu*II ενδονουκλεάσης σε συνθήκες υπερπέψης.

Προηγούμενες εργασίες στο εργαστήριο Ενζυμικής Τεχνολογίας του ΙΜΒΒ {προσωπική επικοινωνία} είχαν δείξει την ιδιότητα της *Pvu*II ενδονουκλεάσης να εμφανίζει σε συνθήκες υπερπέψης (overdigest) ζώνες (εικόνα 8) επιπλέον των αναμενόμενων με βάση την γνωστή αλληλουχία αναγνώρισης. Ενα φαινόμενο γνωστό στα περιοριστικά ένζυμα ως "star" (*) ενεργότητα {Polisky et al., 1975} έχει παρατηρηθεί σε πολλές περιπτώσεις ενδονουκλεασών ως αποτέλεσμα αυξημένης ποσότητας γλυκερόλης, αλάτων η οργανικών διαλυτών στο μίγμα της αντίδρασης. Η (*) ενεργότητα έχει μελετηθεί στην περίπτωση της R.*Pvu*II {Nasri et al., 1987} ως αποτέλεσμα της παρουσίας στο μίγμα της αντίδρασης διμεθυλοσουλφοξειδίου (DMSO) ή/και αιθανόλης. Στις περιπτώσεις που το φαινόμενο της (*) ενεργότητας έχει μελετηθεί {Kalesnikov et al., 1981}, {Nasri & Thomas, 1986} αποδόθηκε σε χαλάρωση (relaxation) της ειδικότητας του ενζύμου με αποτέλεσμα την αναγνώριση στόχων με παραπλήσια αλληλουχία της κανονικής του ενζύμου (εκφύλιση της αλληλουχίας). Η εργασία των Nasri et al. έδειξε ότι η R.*Pvu*II παρουσία DMSO, εκτός της κανονικής αλληλουχίας CAG/CTG, αναγνωρίζει στο πλασμίδιο pBR322 τις χαρακτηρισμένες ως *Pvu*II* αλληλουχίες αναγνώρισης AAG/CTG, GAG/CTG, CNG/CTG, CAN/CTG, CAG/NTG, CAGCNG, CAG/CTC και CAG/CTT. Η *Pvu*II* ενεργότητα αναγνωρίζει και υδρολύει αλληλουχίες οι οποίες διαφέρουν από την κανονική αλληλουχία σε μία μόνο βάση. Ο εκφυλισμός της κύριας αλληλουχίας αναγνώρισης στην περίπτωση αυτή επιτρέπει την αντικατάσταση με οποιαδήποτε βάση κάθε θέσης της αλληλουχίας. Οπως αναφέρθηκε ειδικότερα στην περίπτωση της R.*Pvu*II η (*) ενεργότητα είχε παρατηρηθεί και ως αποτέλεσμα της μεγάλης περίσσειας στην αντίδραση πέψης της ποσότητας του ενζύμου σε σχέση με την ποσότητα του DNA στόχου. Ο εκφυλισμός της αλληλουχίας κατά την υπερπέψη του υποστρώματος είναι περιορισμένος όπως υποδηλώνει η παρουσία μικρού αριθμού επιπλέον θέσεων αναγνώρισης κατά την υδρόλυση μεγάλων σε μέγεθος υποστρωμάτων (εικόνα 8, λ DNA). Η πληροφορία για τις επιπλέον αλληλουχίες αναγνώρισης θα ήταν χρήσιμη σε μετέπειτα στάδια της μελέτης, καθώς γνωρίζοντας τις αλληλεπιδράσεις της πρωτεΐνης με τον υψηλής ειδικότητας στόχο θα μπορούσε να γίνει ανάλυση των αλλαγών που οδηγούν στην αναγνώριση του χαμηλότερης ειδικότητας στόχου.

Ο προσδιορισμός της δεύτερης αυτής αλληλουχίας αναγνώρισης πραγματοποιήθηκε με τροποποίηση της μεθόδου που έχει προταθεί για την εύρεση της κύριας αλληλουχίας αναγνώρισης των περιοριστικών ενζύμων {Gingeras,

Milazzo and Roberts, 1978] και αποτελεί συνδυασμό πειραματικής και θεωρητικής, με την βοήθεια υπολογιστή, προσέγγισης.

DNA από αδενοιό2, πλασμίδιο pBR322 και φάγο λ υδρολύθηκαν σε ιδανικές συνθήκες και σε συνθήκες υπερπέψης (>100X) (κεφάλαιο Δ 3.8.) και αναλύθηκαν σε πηκτώματα αγαρόζης (εικόνες 3α,3β,3γ). Για τα τρία υποστρώματα μεγέθους 35937, 4363 και 48502 βάσεων αντίστοιχα, ακολούθησε προσδιορισμός του μεγέθους των ζωνών και πραγματοποιήθηκε επεξεργασία των δεδομένων αυτών με τον αλγόριθμο που περιγράφεται στον πίνακα 5. Για την εφαρμογή του αλγορίθμου γράφτηκαν τα προγράμματα τα οποία παρατίθενται στο προσάρτημα III.



Εικόνα 8. Ανάλυση αντιδράσεων υπερπέψης ολικού DNA *από α) 1μg αδενοιού 2 β) 1μg pBR322 και γ) 1μg φάγου λ (6 ώρες με 240-160-80-0 μονάδες ενζύμου) σε πηκτώματα αγαρόζης 1%. Y αντίδραση υπερπέψης και K πέψη σε ιδανικές συνθήκες για την R.PvuII.*

Η πορεία επιλογής των πιθανών αλληλουχιών αναγνώρισης παρουσιάζεται στον πίνακα 6. Η αλληλουχία AAGCTT είναι η μοναδική ικανή να δώσει τα πρότυπα πέψης που παρατηρούνται για τα τρία υποστρώματα και είναι όμοια με αυτή του ενζύμου *Hind*III. Σε αντίθεση με τα αποτελέσματα από την παρουσία DMSO στην αντίδραση, είναι ενδιαφέρον ότι, σε σχέση με την "φυσιολογική" αλληλουχία αναγνώρισης της R.*Pvu*II CAG/CTG, η δευτερεύουσα αλληλουχία διαφοροποιείται στις δύο εξωτερικές βάσεις (C->A, G->T), ικανοποιώντας έτσι το παλίνδρομο της αλληλουχίας. Για την επαλήθευση του αποτελέσματος έγινε σύγκριση του προτύπου πέψης λ-DNA με μίγμα R.*Pvu*II και R.*Hind*III, με αυτό από την υπερπέψη με R.*Pvu*II (εικόνα 9).

Πίνακας 5. Αλγόριθμος για την εύρεση της (*) αλληλουχίας αναγνώρισης περιοριστικών ενζύμων.

1) Δημιουργία καταλόγου με όλα τα διαφορετικά εξανουκλεοτίδια[1] που περιέχονται στην αλληλουχία ενός από τα επιλεχθέντα υποστρώματα. ➥
2) Προσομοίωση και αποθήκευση των προτύπων πέψης από καθένα (ή συνδιασμό) από τα εξανουκλεοτίδια του προηγούμενου καταλόγου, σε συνδιασμό με την φυσιολογική αλληλουχία αναγνώρισης, για το συγκεκριμένο υπόστρωμα. ➥
3) Ελεγχος για συμφωνία μεταξύ των θεωρητικών προτύπων από το βήμα 2 και του αντίστοιχου πειραματικού και επιλογή των εξανουκλεοτιδίων των οποίων τα πρότυπα πέψης συμφωνούν, στα πλαίσια της ακρίβειας προσδιορισμού του μεγέθους των ζωνών από τα πηκτώματα αγαρόζης. ➥
4) Επιστροφή στο βήμα 2 με τα επιλεγμένα εξανουκλεοτίδια και θεωρητική πέψη του DNA δεύτερου (τρίτου κ.ο.κ...) υποστρώματος. ➡
5) Τα βήματα 2-4 επαναλσμβάνονται μέχρι να εξαντληθούν όλες οι διαθέσιμες αλληλουχίες υποστρωμάτων ή να βρεθεί μία μοναδική αλληλουχία αναγνώρισης που να ικανοποιεί όλα τα διαθέσιμα πρότυπα.

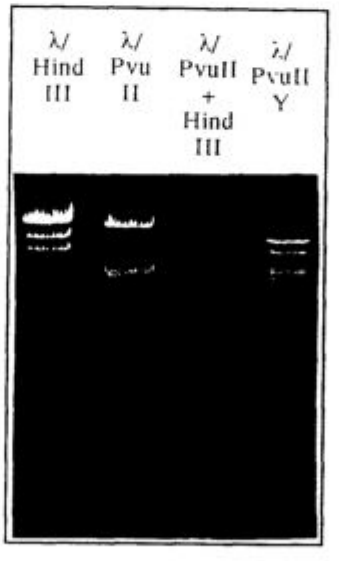
[1] Εφαρμογή του ίδιου αλγορίθμου μπορεί να γίνει για αλληλουχίες αναγνώρισης διαφορετικού μεγέθους (π.χ. πεντανουκλεοτίδια) ή για περισσότερες της μίας * αλληλουχίες αναγνώρισης.

✍

Πίνακας 6. Πορεία επιλογής εξανουκλεοτιδίων στα βήματα της διαδικασίας εύρεσης της δευτερεύουσας αλληλουχίας αναγνώρισης της R.*Pvu*II.

Σύνολο δυνατών εξανουκλεοτιδίων 4^6=4096	
pBR322	2596 διαφορετικά εξανουκλεοτίδια.
pBR322	156 εξανουκλεοτίδια με πρότυπο πέψης σύμφωνο με το πειραματικό.
Λάμδα DNA	3 εξανουκλεοτίδια με πρότυπο πέψης σύμφωνο με το πειραματικό.
Αδενοιός	1 εξανουκλεοτίδιο με αλληλουχία AAGCTT.

✍



Εικόνα 9. Σύγκριση πέψης λ-DNA από PvuII/Hind III με αυτό από υπερπέψη με PvuII. Y : αντίδραση υπερπέψης.

Β 3. Απομόνωση και καθαρισμός του ενζύμου.

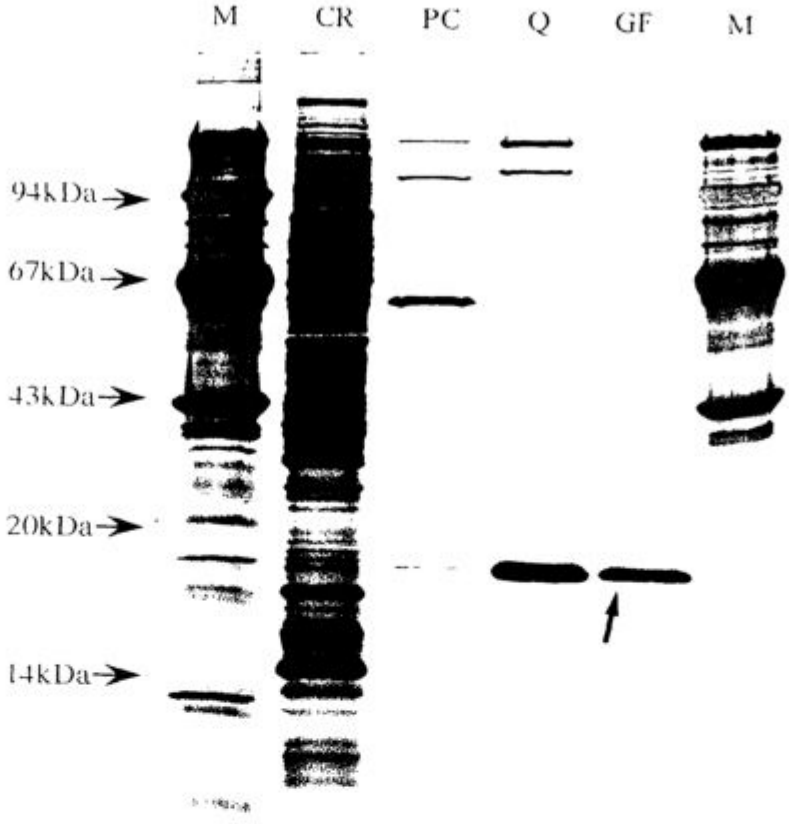
Στόχος αυτού του τμήματος της εργασίας υπήρξε η παραγωγή της πρωτεΐνης σε ποσότητες ικανές για τα πειράματα κρυστάλλωσης και καθαρότητας που να εξασφαλίζει την ομοιογένεια του πρωτεϊνικού δείγματος που θα χρησιμοποιηθεί για τα πειράματα αυτά. Σαν πηγή για την παραγωγή της πρωτεΐνης χρησιμοποιήθηκε το *E.coli* στέλεχος HB101 μετασχηματισμένο με το πλασμίδιο p*Pvu*RM3.4. Η μεταγραφή των γονιδίων του *Pvu*II συστήματος στον κλώνο αυτό είναι κάτω από τον έλεγχο των υποκινητών των γονιδίων στον πατρικό οργανισμό *Proteus vulgaris*. Μικρή αύξηση στην παραγωγή της πρωτεΐνης στην περίπτωση αυτή επιτυγχάνεται λόγω του μεγάλου αριθμού αντιγράφων του πλασμιδίου pUC18 που αποτελεί τον φορέα της p*Pvu*RM3.4 κατασκευής.

Για την παραγωγή της πρωτεΐνης σε μεγάλη κλίμακα πραγματοποιήθηκε μια σειρά καλλιεργειών σε ζυμωτήρα 30lt του εργαστηρίου Ενζυμικής Τεχνολογίας του ΙΜΒΒ. Για τις καλλιέργειες αυτές χρησιμοποιήθηκαν συνήθη θρεπτικά υλικά για την καλλιέργεια *E.coli* στελεχών, σε ενισχυμένες δόσεις (βλ. κεφάλαιο Δ 3.1.). Η παραγωγή βακτηριακής πάστας από τις καλλιέργειες αυτές υπήρξε της τάξης των 200gr κατά μέσον όρο και η αποθήκευση της έγινε στους -70° C.

Για την λύση των κυττάρων χρησιμοποιήθηκαν οι μέθοδοι French-Press, υπερηχοβόλησης καθώς και ενζυματικής λύσης με λυσοζύμη (βλ. κεφάλαιο Δ 3.2., {Cull & McHenry, 1990}). Αποτελεσματικότερη μέθοδος αποδείχθηκε η ηπιότερη μέθοδος : αυτή της ενζυματικής λύσης, εξασφαλίζοντας τις μικρότερες απώλειες πρωτεΐνης. Η συγκεκριμένη μεθοδολογία ελαχιστοποιώντας τον χρόνο από την χρονική στιγμή της λύσης των κυττάρων μέχρι το φόρτωμα του πρωτεϊνικού εκχυλίσματος στην πρώτη χρωματογραφική στήλη μείωσε την δράση των πρωτεασών που βρίθουν στο ακατέργαστο εκχύλισμα βακτηριακών κυττάρων.

Οι αρχικοί καθαρισμοί της πρωτεΐνης στόχευσαν στην δημιουργία ενός πρώτου πρωτοκόλλου ικανού να δώσει πρωτεΐνη υψηλής καθαρότητας ανεξαρτήτως της ποσότητας της τελικά καθαρής πρωτεΐνης. Σταδιακές βελτιώσεις ενός πρώτου σχήματος καθαρισμού με την αφαίρεση ενός σταδίου κλασμάτωσης με θειικό αμμώνιο και την αντικατάσταση της μηχανικής λύσης των κυττάρων με ενζυμική οδήγησαν στο τελικό σχήμα καθαρισμού που παρουσιάζεται στον πίνακα 7. Πρωτεινικά δείγματα από το σχήμα αυτό έδωσαν τους κρυστάλλους για την συλλογή των δεδομένων που χρησιμοποιήθηκαν για την επίλυση της δομής. Στο κεφάλαιο Δ 3.4. παρατίθενται λεπτομέρειες του πρωτοκόλλου καθαρισμού της R.*Pvu*II.

Η ηλεκτροφορητική καθαρότητα της πρωτεΐνης στα στάδια της απομόνωσης της πρωτεΐνης παρουσιάζεται στην εικόνα 10 και τα στατιστικά ενός



Εικόνα 10. Ανάλυση σε αποδιατακτικό πήκτωμα ακρυλαμίδης 12% δειγμάτων από τα στάδια του καθαρισμού της PvuII ενδονουκλεάσης. *Cr : εκχύλισμα κυττάρων (2.70μgr), PC : (Κολώνα Φωσφοκυτταρίνης) (0.34μgr), Q : (Q-Sepharose) (1.05μgr), και GF : στήλη μοριακής διήθησης Sephadex75 (0.32μgr) και M : μάρτυρας μοριακού βάρους. Με βέλος η ζώνη της R.PvuII και τα πρότυπα μοριακά βάρη. Χρώση του πηκτώματος έγινε με νιτρικό άργυρο.*

καθαρισμού από 80gr βακτηριακής πάστας παρουσιάζονται στον πίνακα 8. Το μοριακό βάρος της R.*Pvu*II όπως προσδιορίζεται σε αποδιατακτικό πήκτωμα ακρυλαμίδης είναι κατά προσέγγιση 18000Da και είναι σύμφωνο με αυτό που προβλέπεται από την ανάλυση της νουκλεοτιδικής αλληλουχίας (κεφάλαιο Β1.). Στα πειράματα μοριακής διήθησης η ενεργότητα της ενδονουκλεάσης εκλύεται σε φαινόμενο μοριακό βάρος 40kDa {Tao and Blumenthal, 1992}, {δεδομένα που δεν παρουσιάζονται} που είναι σύμφωνο με τα δεδομένα για την παρουσία της R.*Pvu*II στο διάλυμα ως διμερές {Gingeras et al., 1981}. Η καθαρότητα του

Πίνακας 7. Συνοπτική περιγραφή του πρωτοκόλλου καθαρισμού της *Pvu*II ενδονουκλεάσης. Το στάδιο 6 είναι προαιρετικό για την παραγωγή πρωτείνης για κρυσταλλώσεις. Λεπτομέρειες του πρωτοκόλλου παρατίθενται στο κεφάλαιο Δ 3.4.

1) Ενζυματική λύση με λυσοζύμη.
2) Πέψη νουκλεικών οξέων με νουκλεάση.
3) Φυγοκέντριση στα 19000 g (Φυγ. Sigma 2K15)
4) Κολώνα φωσφοκυτταρίνης 300ml, διαβάθμιση KCl 0.1-0.6 M, pH 7.4.
5) Κολώνα ιοντοανταλλαγής MonoQ 8ml (FPLC) η Q-Sepharose, διαβάθμιση KCl 0.01-0.3M, pH 7.4.
6) Κολώνα μοριακής διήθησης Sephadex G75, pH 7.4.

Πίνακας 8. Στατιστικά ενός καθαρισμού της R.*Pvu*I. Η ενεργότητα στο δείγμα μετά την λύση των κυττάρων δεν μπορεί να προσδιορισθεί με ακρίβεια λόγω της παρουσίας μη ειδικών νουκλεασών, έτσι δεν υπάρχει εκτίμηση για τις απώλειες κατά το στάδιο της στήλης Φωσφοκυτταρίνης.

Βήμα	Ποσότητα Πρωτείνης (mg)	Συνολική ενεργότητα (u)*	Ειδική ενεργότητα (u/mg)	Ανάκτηση (%)	Καθαρισμός (φορές)
Εκχύλιση	1125	-	-	-	-
Φωσφο-κυτταρίνη	33.12	240000	7246	100	1
Q-Sepharose	1.47	210000	142857	87.5	19.7
Sephadex 75	0.57	135000	236842	56.2	32.7

* Η μονάδα ενεργότητας (u) του ενζύμου ορίζεται ως η απαραίτητη ποσότητα του ενζύμου για την πλήρη πέψη 1μgr λ-DNA σε 1 ώρα στους 37° C.

πρωτεϊνικού δείγματος από τους καθαρισμούς ελεγχόταν σε αποδιατακτικά πηκτώματα ακρυλαμίδης με χρώση νιτρικού αργύρου (βλ. κεφάλαιο Δ 3.6.). Επιλογή των κλασμάτων του ενζύμου από τις κολώνες χρωματογραφίας έγινε με συνδυασμό πηκτωμάτων ακρυλαμίδης και παράλληλη παρακολούθηση της ενεργότητας με ανάλυση των αντιδράσεων πέψης λ-DNA με δείγματα από τα κλάσματα, σε πηκτώματα αγαρόζης (βλ. κεφάλαιο Δ 3.5.).

Υστερα από τον προσδιορισμό των συνθηκών κρυστάλλωσης της πρωτεΐνης που περιγράφεται σε επόμενο κεφάλαιο, πραγματοποιήθηκαν πειράματα κρυστάλλωσης με δείγματα πρωτεΐνης από στάδια του καθαρισμού πρίν από το τελικό. Αποδείχθηκε ότι η κρυστάλλωση της πρωτεΐνης στις συνθήκες τις οποίες χρησιμοποιήσαμε δεν επηρεάζεται από την παρουσία άλλων πρωτεϊνών που ακολουθούν την *Pvu*II ενδονουκλεάση μετά το στάδιο της στήλης ιοντοανταλλαγής MonoQ/Q-FastFlow (εικόνα 10, Q). Καθαρό δείγμα πρωτεΐνης

μετά την στήλη μοριακής διήθησης σε συγκέντρωση κατάλληλη για πειράματα κρυστάλλωσης (1-20 mg/ml) παρουσιάζει σε διάστημα 2-3 ημερών έντονο το πρόβλημα σχηματισμού συσσωματωμάτων (aggregation) και μη αντιστρεπτής ιζηματοποίησης με αποτέλεσμα το δείγμα να καθίσταται μη χρησιμοποιήσιμο για πειράματα κρυστάλλωσης. Πλεονέκτημα της μη ομογενώς καθαρής πρωτεΐνης από το στάδιο Q-Sepharose/MonoQ αποτελεί το γεγονός ότι δεν εμφανίζει το φαινόμενο της συσσωμάτωσης, με αποτέλεσμα την αύξηση του χρόνου ζωής της πρωτεΐνης στις συνθήκες αποθήκευσης, γεγονός που αποτέλεσε παράγοντα καθοριστικό για την επιτυχία των κρυσταλλώσεων.

Σαν αποτέλεσμα το στάδιο μοριακής διήθησης δεν πραγματοποιήθηκε σε αρκετές περιπτώσεις καθαρισμών.

Β 4. Κρυστάλλωση της *Pvu*II ενδονουκλεάσης.

Σε δύο στάδια μπορούν να διακριθούν τα πειράματα κρυστάλλωσης της *Pvu*II ενδονουκλεάσης. Στην πρώτη φάση έγινε ανίχνευση των συνθηκών κρυστάλλωσης και χαρακτηρισμός των κρυστάλλων που προέκυψαν {Athanasiadis & Kokkinidis, 1991} ενώ στην δεύτερη βελτιστοποίηση των κρυστάλλων και παραγωγή τους για κρυσταλλογραφικές μετρήσεις σε μέση και ψηλή διακριτικότητα. Για την κρυστάλλωση της *Pvu*II ενδονουκλεάσης χρησιμοποιήθηκε εκτεταμένα η μέθοδος της διάχυσης ατμών (vapour diffusion) είτε με την τεχνική "hanging drops" (βλ. κεφάλαιο Δ 4.1.) είτε με "sitting drops" {McPherson, 1990}, {Gilliland & Davies, 1990}. Αναζήτηση των συνθηκών κρυστάλλωσης της πρωτεΐνης έγινε με την χρήση θειικού αμμωνίου (AS), (καθώς και οργανικών διαλυτών EtOH, MetOH, PEG6000, PEG4000) ως παράγοντα καθίζησης σε "hanging drop" πειράματα όπου ελέγχθηκαν περιοχές του pH (3 - 9) όπως και της συγκέντρωσης θειικού αμμωνίου (20-50% κεκορεσμένου διαλύματος AS). Μικρών διαστάσεων (0.02x0.02x0.1mm) βελονοειδείς κρύσταλλοι της R.*Pvu*II εμφανίστηκαν αρχικά σε συνθήκες όπου η τελική σύσταση του πρωτεϊνικού διαλύματος ήταν 47% σε θειικό αμμώνιο και σε pH 4.8. Στον πίνακα 9 αναγράφονται οι συνθήκες κρυστάλλωσης της πρωτεΐνης ύστερα από βελτιστοποίηση των αρχικών συνθηκών. Τυπικά, ραβδόμορφοι κρύσταλλοι (εικόνα 11α) εμφανίζονται σε δύο μέρες και συνεχίζουν την ανάπτυξη τους για διάστημα δύο εβδομάδων καταλήγοντας σε μέγεθος 0.3x0.2x1.0mm. Στις ίδιες συνθήκες κρυστάλλωσης και πολλές φορές στο ίδιο πείραμα παρατηρείται η εμφάνιση κρυστάλλων σφηνοειδούς μορφής (εικόνα 11β). Ατέλειες των κρυστάλλων εμφανίζονται στην διάρκεια της ανάπτυξης τους υπο μορφήν κρυσταλλικών προεξοχών που αναπτύσονται από το κέντρο του μητρικού κρυστάλλου (εικόνα 11γ). Μικρό ποσοστό των κρυστάλλων φτάνει στις τελικές του διαστάσεις χωρίς την εμφάνιση προεξοχών. Προσεκτική αφαίρεση των προεξοχών, ελαφρά διαλυτοποίηση της επιφάνειας του κρυστάλλου και μεταφορά του σε νέο μητρικό διάλυμα επιτρέπει την αξιοποίηση των ατελών κρυστάλλων (βλ. κεφάλαιο Δ 4.2.). Κρύσταλλοι της *Pvu*II ενδονουκλεάσης μπορούν να παραχθούν με την χρήση και άλλων παραγόντων καθίζησης όπως το $MgSO_4$ και Li_2SO_4, μολονότι χειρότερης ποιότητας.

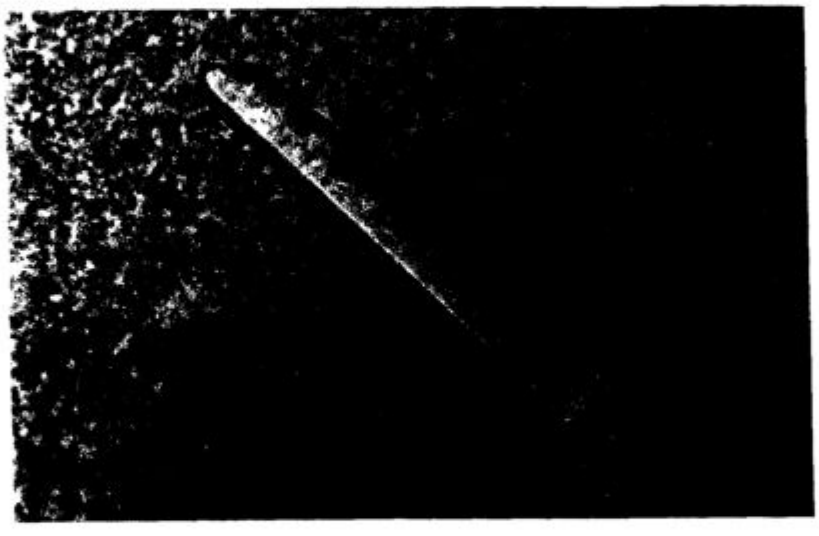
Επαναδιάλυση κρυστάλλων της R.*Pvu*II που παρέμειναν στο μητρικό διάλυμα για χρονικό διάστημα μεγαλύτερο από ένα χρόνο έδωσε πρωτεϊνικό δείγμα που συμφωνεί στο μοριακό βάρος με αυτό της πρωτεΐνης και διατηρεί την ενεργότητα του (εικόνες 12α,12β).

Πίνακας 9. Συνθήκες κρυστάλλωσης της *Pvu*II ενδονουκλεάσης σε πειράματα "hanging drops".

Πρωτείνη (αποθήκευση)	Αρχικές συνθήκες (σταγόνα)	Συνθήκες εξισορρόπησης (δεξαμενή)
5-15mg/ml σε 20mM Tris.Cl pH 7.4, 20mM KCl	10-20% AS, 100 mM Tris.Cl pH 5.0-5.2 2.5-7.5mg/ml R.*Pvu*II	44% AS, 100mM Tris.Cl pH 5.0-5.2

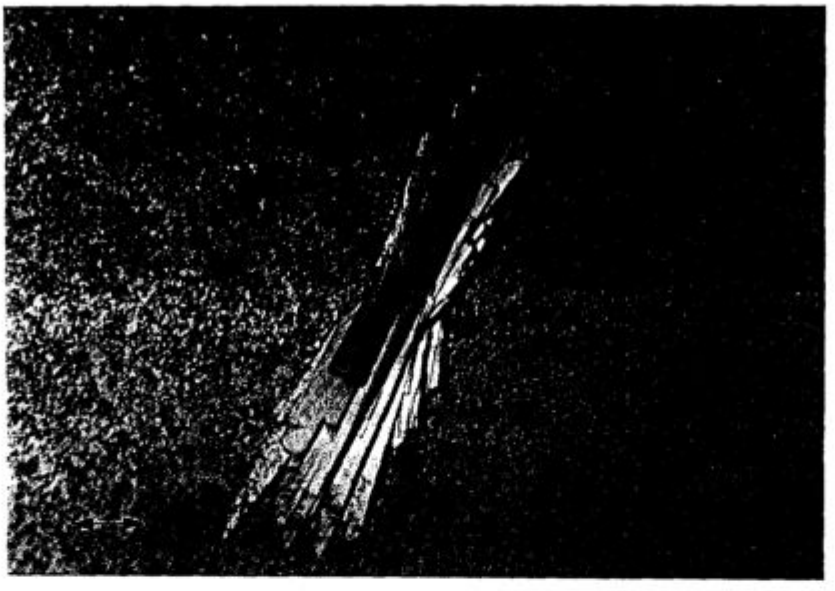
Εικόνα 11. Μορφολογία και διαστάσεις κρυστάλλων της R.PvuII. *Κρύσταλοι ανεπτυγμένοι στις συνθήκες του πίνακα 9 σε μεγένθυση 50X (Το σύμβολο στις φωτογραφίες αντιστοιχεί σε 0.20mm πραγματικό μέγεθος).*

α) τυπικός ραβδόμορφος κρύσταλλος.

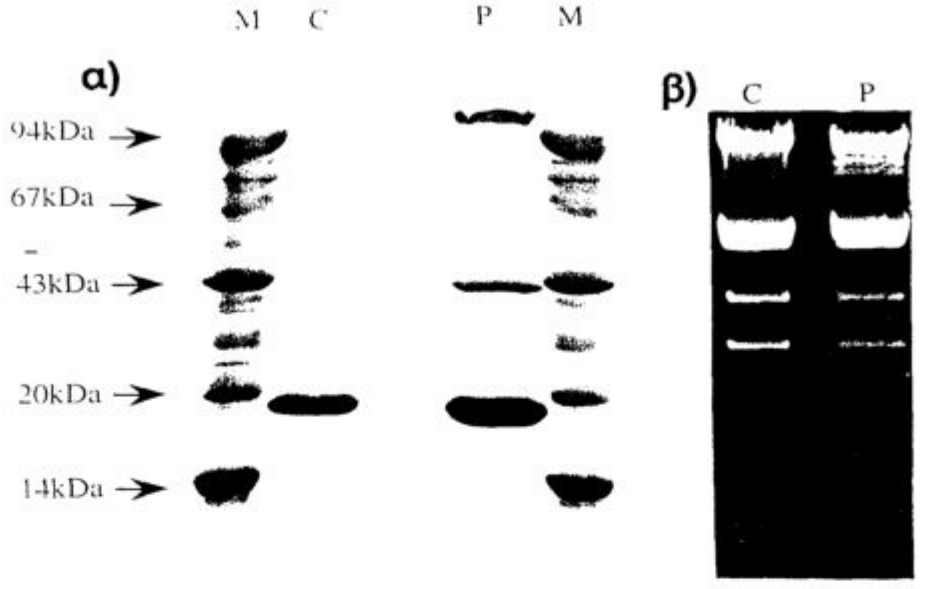


β) σφηνοειδής και ραβδόμορφος κρύσταλλος στην ίδια σταγόνα.

γ) χαρακτηριστικές ατέλειες κρυστάλλων της R.*Pvu*II.



❖



Εικόνα 12. Πρωτείνη από επαναδιάλυση κρυστάλλων της R.PvuII. *α) Ανάλυση σε αποδιατακτικό πήκτωμα ακρυλαμίδης 12% (χρώση Coomasie Briliant Blue) M : Μάρτυρας μοριακού βάρους C : Πρωτείνη από επαναδιάλυση κρυστάλλων P: Πρωτείνη που χρησιμοποιήθηκε για κρυστάλλωση. β) Πέψη 1μg λ-DNA από κρυσταλλική πρωτεΐνη (C) και καθαρή πρωτεΐνη (P). Η επαναδιάλυση των κρυστάλλων πραγματοποιήθηκε ώς εξής: Κρύσταλλοι της R.PvuII μαζί με μικρή ποσότητα μητρικού υγρού τοποθετήθηκαν σε σωλήνα eppendorf και ακολούθησαν τέσσερις κύκλοι πλυσίματος (προσθήκη 1 ml νερού, φυγοκέντριση 1', απομάκρυνση υπερκειμένου και εκ νέου προσθήκη 1ml νερού) για την απομάκρύνση του μητρικού υγρού. Στην συνέχεια οι κρύσταλλοι αφέθηκαν να επαναδιαλυθούν σε 20mM Tris-HCl pH7.4 για τουλάχιστον 12 ώρες πρίν χρησιμοποιηθεί το δείγμα.*

Β 5. Συλλογή κρυσταλλογραφικών δεδομένων.

Β 5.1. Αρχικός κρυσταλλογραφικός χαρακτηρισμός των κρυστάλλων της ενδονουκλεάσης *Pvu*II.

Ραβδοειδείς κρύσταλλοι διαστάσεων 0.5mm x 0.3mm x 0.2mm κλεισμένοι σε γυάλινα τριχοειδή σωληνάρια (capillaries) εκτέθηκαν σε ακτίνες-X σε ένα CAD4 περιθλασίμετρο της εταιρίας Enraf-Nonius. Η μέτρηση έγινε σε θερμοκρασία δωματίου και οι συνθήκες λειτουργίας της πηγής ήταν 40kV και 32mA. Συνολικά καταγράφηκαν 25 ανακλάσεις μεταξύ 10Å και 3Å. Με βάση τις ανακλάσεις αυτές προσδιορίστηκαν οι παράμετροι της στοιχειώδους κυψελίδας. Ακολούθησε συλλογή δεδομένων μέχρι τα 5Å διακριτικότητα προκειμένου να προσδιοριστούν οι συστηματικές απουσίες ανακλάσεων. Τα αποτελέσματα συνοψίζονται στον πίνακα 10. Το V_M όπως ορίστηκε από τον Matthews {Mathews, 1968} υπολογίστηκε για διαφορετικές τιμές αριθμού μορίων ανά ασύμμετρη μονάδα της πρωτεΐνης δίνοντας τυπική τιμή για πρωτεΐνη στην περίπτωση δύο μορίων ανά ασύμμετρη μονάδα (επιτρεπτές τιμές 1.68-3.53 $Å^3$/dalton). Οι τιμές αυτές παρουσιάζονται στον πίνακα 11 μαζί με τα αντίστοιχα ποσοστά περιεκτικότητας της ασύμμετρης μονάδας σε διαλύτη. Το ποσοστό περιεκτικότητας της ασύμμετρης μονάδας σε διαλύτη υπολογίζεται κατά προσέγγιση με βάση στατιστική ανάλυση δείγματος 116 διαφορετικών κρυσταλλομένων πρωτεϊνών όπου το V_M κυμαίνεται μεταξύ 1.68-3.53 $Å^3$ ενώ αντίστοιχα η περιεκτικότητα σε διαλύτη ανάμεσα σε 27%-65% {Mathews, 1968}.

Απόσβεση των ανακλάσεων παρατηρήθηκε κατά μήκος των αξόνων για h=2n+1 και k=2n+1, αντίθετα στον άξονα l η παρουσία της ανάκλασης [005] στην χαμηλή διακριτικότητα μας έδειξε ότι δεν έχουμε αποσβέσεις. Ολες οι μετρήσεις κρυστάλλων που ακολούθησαν σε ψηλότερη διακριτικότητα υπήρξαν σύμφωνες σε ότι αφορά της συστηματικές αποσβέσεις κατά μήκος των αξόνων.

Πίνακας 10. Παράμετροι της στοιχειώδους κυψελίδας κρυστάλλων της *Pvu*II ενδονουκλεάσης.

a=84.20Å, b=106.20Å, c=46.9Å, α=β=γ=90°
Χωροομάδα συμμετρίας **$P2_12_12$** (No. 18: {International Tables for X-ray Crystallography, 1989}).

Η παρουσία δύο μορίων στην ασύμμετρη μονάδα υποδηλώνει με μεγάλη πιθανότητα την παρουσία ενός μή κρυσταλλογραφικού άξονα τάξεως-2 (2-fold) ο οποίος όπως θα δούμε στην συνέχεια της εργασίας συμπίπτει με τον μοριακό άξονα που σχηματίζει το διμερές στο διάλυμα.

Οι κρύσταλλοι της *Pvu*II ενδονουκλεάσης έδειξαν χαμηλή ευπάθεια στην ακτινοβόλιση με ακτίνες-X και ψηλό βαθμό ισομορφισμού (Ολοι οι κρύσταλλοι της φυσικής πρωτεΐνης που μετρήθηκαν αποκλίνουν λιγότερο από 1% στις παραμέτρους της στοιχειώδους κυψελίδας).

Πίνακας 11. Τιμές V_M για διαφορετικούς αριθμούς μονομερών ανά ασύμμετρη μονάδα.
$V_M=V_C/(MW^*N)$ {Blundell & Johnosn, 1976}, {Mathews, 1968} όπου V_C=Ογκος της ασύμμετρης μονάδας, MW= Μοριακό βάρος της πρωτείνης και N ο αριθμός μορίων ανά ασύμμετρη μονάδα. (*Pvu*II MW =18300 Da για το μονομερές). Το ποσοστό του όγκου της πρωτείνης στην ασύμμετρη μονάδα δίνεται από τον τύπο $V_P=A_V/(N_A \cdot D_P \cdot V_M)$ όπου $A_V=10Å^3$ ο μέσος ατομικός όγκος, $N_A=6.023$ είναι η σταθερά Avogadro, και $D_P=1.35gr/cm^3$ η μέση πυκνότητα πρωτεΐνης. (Η περιεκτικότητα σε διαλύτη 1-Vp)

Μονομερή/ ασύμμετρη μονάδα	V_M A^3/dalton	Περιεκτικότητα σε διαλύτη. (%)
1	5.78	78.75
2	***2.89***	***57.45***
3	1.92	36.03
4	1.44	14.7

Β 5.2. Συλλογή δεδομένων για την φυσική πρωτεΐνη.

Τρία σύνολα δεδομένων από διαφορετικά είδη πηγών ακτίνων X, συλλέχθηκαν συνολικά για την φυσική (native) πρωτεΐνη. Στατιστικά των μετρήσεων αυτών παρουσιάζονται στον πίνακα 12. Από τις σειρές αυτές τα δεδομένα τα οποία συλλέχθηκαν με ακτινοβολία σύγχροτρον είναι αυτά που τελικά χρησιμοποιήθηκαν για τον προσδιορισμό των φάσεων της πρωτεΐνης ενώ τα υπόλοιπα χρησιμοποιήθηκαν σε μερικές περιπτώσεις για τον χαρακτηρισμό παραγώγων των κρυστάλλων με βαριά άτομα. Η επεξεργασία των δεδομένων που συλλέχθηκαν με δίσκο ειδώλου (image plate) από ακτινοβολία σύγχροτρον στον δακτύλιο DORIS II του DESY (Deutsches Electronen SYnchrotron) έγινε με το πακέτο λογισμικού MOSFLM(v50) {Leslie, Brick and Wonacott, 1986}. Για την αντίστοιχη επεξεργασία των δεδομένων από ανιχνευτή περιοχής (area detector) χρησιμοποιήθηκε το πακέτο προγραμμάτων XDS {Kabsch, 1988}. Τέλος για τα δεδομένα του περιθλασίμετρου χρησιμοποιήθηκαν προγράμματα που συνοδεύουν CAD4 περιθλασίμετρο και του CCP4 πακέτου κρυσταλλογραφικών

προγραμμάτων, τροποποιημένα για να ικανοποιήσουν τις ανάγκες του εργαστηρίου.

Πίνακας 12. Στατιστικά συλλογής δεδομένων για την φυσική πρωτεΐνη. Με έντονους χαρακτήρες τα τρία σύνολα δεδομένων (data sets) όπως προέκυψαν από την συγχώνευση των επιμέρους μετρήσεων. NatM είναι το σύνολο των δεδομένων στο οποίο βασίστηκε η επίλυση της δομής.

Σειρά	NatH	NatL	**NatM**	Nat1	Nat1a	Nat2	Nat3	**NatMH**	**Native**
Πηγή			**X11**					**Π.Α.†**	**CuKa**
Μήκος κύματος			**0.92**					**1.54**	**1.54**
Ανιχνευτής			**MAR**					**Xen.**	**CAD4**
Αρ. κρυστάλλων			**1**					**3**	**1**
Ημερομηνία			**10/12/92**					**13/12/91**	**16/8/93**
Μέγιστη Διακριτικότητα (Å)	2.3	3.7	**2.4**	3.65	3.65	2.58	2.58	**3.50**	**6.00**
a (Å)*			**84.65**		84.07	84.04	84.06	**84.01**	**84.24**
b (Å)*			**106.46**		106.15	105.99	105.99	**105.96**	**105.89**
c (Å)*			**47.00**		46.97	46.94	46.92	**46.92**	**46.89**
I>3σI (%)[1]			**79.2**					**63.0**	**51.7**
I<1σI (%)[1]			**5.4**					**8.6**	**23.5**
Αρ. Ανακλάσεων	16870		**16315**	2981	2547	6331	7145	**8490**	**1199**
R_{sym}	7.4		**7.6**	10.0	12.6	8.9	14.0	**13.2**	**16.1**
Πληρότητα%	86.10		**94.30**	60.3	52.2	46.6	52.6	**94.5**	**100**
$\langle F\rangle/\langle\sigma(F)\rangle$**			**20.68**	8.24	9.19	5.19	5.36	**6.80**	**7.17**

[1] I η ένταση μιάς ανάκλασης [hkl]
* Οι παράμετροι της στοιχείωδους κυψελίδας δεν παρέχονται όπου υπάρχουν άλλες μετρήσεις από τον ίδιο κρύσταλλο.
** F το πλάτος συντελεστή δομής και σ(*F*) η τυπική απόκλιση για το πλάτος του συντελεστή δομής.
X11 : Γραμμή ακτινοβολίας ακτίνων-X στον δακτύλιο DORIS II του DESY.
† : Περιστρεφόμενη άνοδος GX18.

$$R_{symm}=\frac{\sum_h (I_h-\overline{I_h})}{\sum_h \overline{I_h}}$$

I_h η ένταση της ανάκλασης [hkl] και $\langle I_h\rangle$ η μέση ένταση των ισοδύναμων ανακλάσεων της [hkl].

Η μέτρηση του κυρίου συνόλου δεδομένων NatM έγινε σε δύο στάδια, πρώτα μαζεύτηκε ένα σύνολο δεδομένων σε ψηλή διακριτικότητα (2.3Å) και καθώς για τις χαμηλής διακριτικότητας ανακλάσεις εμφανιζόταν υπερφόρτωση (overloading) του δίσκου ειδώλου συλλέχθηκε ένα δεύτερο σύνολο δεδομένων μέχρι τα 3.7Å μειώνοντας τον χρόνο έκθεσης του κρυστάλλου και κατά συνέπεια την ένταση των ανακλάσεων. Τα δύο σύνολα δεδομένων συλλέχθηκαν από τον ίδιο κρύσταλλο, σε θερμοκρασία δωματίου, ενώ ο χρόνος από την πρώτη έκθεση του κρυστάλλου στις ακτίνες X μέχρι την εκκίνηση της συλλογής του χαμηλής

διακριτικότητας συνόλου ήταν περίπου 7 ώρες για να ολοκληρωθεί σε συνολικό χρόνο έκθεσης 10 ωρών. Για το υψηλής διακριτικότητας σύνολο δεδομένων οι συνθήκες λειτουργίας της πηγής κατα την εκκίνηση της μέτρησης ήταν 24mA και συχνότητα 80kHz, η έκθεση του κρυστάλλου για την συλλογή κάθε ειδώλου ισοδυναμούσε σε 5000 κρούσεις και κατά προσέγγιση σε χρόνο 50", για το χαμηλής διακριτικότητας σύνολο η ένταση της δέσμης ήταν 19mA και η συχνότητα 44kHz. Για να επιτευχθεί λογικός χρόνος έκθεσης καθώς η πηγή ήταν ιδιαίτερα ισχυρή περεμβλήθηκαν φύλλα αλουμινίου και η έκθεση έγινε για 700 κρούσεις (30"). Η γωνία ταλάντωσης (oscillation angle) ήταν 1° για υψηλής διακριτικότητας σύνολο και 2° για το χαμηλής. Στην διάρκεια της μέτρησης η μορφολογική κατάσταση του κρυστάλλου παρακολουθώταν με ειδικά τοποθετημένη οθώνη χωρίς να διαπιστοθούν ίχνη αλλίωσης του κρυστάλλου απο την έκθεση στην ακτινοβολία. Ο κρύσταλλος διαστάσεων 0.5mm x 0.2mm x 0.2mm προερχόταν από βελτίωση αρχικού κρυστάλλου με την διαδικασία μάκρο-σποράς (βλ. κεφάλαιο Δ 4.2.). Οι δύο σειρές συγχωνεύτηκαν στην σειρά NatM αφού έγινε απομάκρυνση των ισχυρών ανακλάσεων από τη σειρά της υψηλής διακριτικότητας όπου εμφανιζόταν υπερφόρτωση του δίσκου ειδώλου. Η σειρά NatM χρησιμοποιήθηκε στις περισσότερες περιπτώσεις για τον προσδιορισμό των θέσεων των βαρειών ατόμων στα παράγωγα όπως και των φάσεων της φυσικής πρωτεΐνης και την κατασκευή των χαρτών ηλεκτρονιακής πυκνότητας της πρωτεΐνης.

Η επεξεργασία των δεδομένων της φυσικής πρωτεΐνης έγινε με την διαδικασία που περιγράφεται στο κεφάλαιο Δ 5.1 και υπήρξε ίδια για όλα τα σύνολα δεδομένων για τα οποία πραγματοποιήθηκαν μετρήσεις στον X11 σταθμό ακτινοβολίας σύγχροτρον.

Συγχώνευση 4 συνόλων δεδομένων από 3 διαφορετικούς κρυστάλλους έδωσε το σύνολο δεδομένων NatMH στα 3.5Å. Οι μετρήσεις για το σύνολο αυτό πραγματοποιήθηκαν σε πηγή ακτίνων-X περιστρεφόμενης ανόδου με την χρήση ανιχνευτή περιοχής Xentronics. Το σύνολο αυτό δεδομένων χρησιμοποιήθηκε στα αρχικά στάδια της αναζήτησης παραγώγων βαρειών ατόμων.

Σε όλες τις περιπτώσεις η επεξεργασία των δεδομένων οδήγησε στην δημιουργία ενός αρχείου ανακλάσεων τύπου MTZ ή LCF κατάλληλη για χρήση με το κρυσταλλογραφικό πακέτο προγραμμάτων CCP4 {CCP4, 1979}.

✂

Β 5.3. Προετοιμασία και συλλογή δεδομένων των παραγώγων βαρειών ατόμων των κρυστάλλων της R.*Pvu*II.

Η επιλογή των ενώσεων βαρειών ατόμων είναι συνήθως διαδικασία δοκιμής και σφάλματος {Leslie, 1991}. Η γνώση της πρωτεϊνικής αλληλουχίας της *Pvu*II έδωσε λίγα στοιχεία για την αρχική επιλογή ενώσεων για την παραγώγιση. Η ειδικότητα ενώσεων του αργύρου (Ag) για κατάλοιπα His υπέδειξε μια πρώτη επιλογή με δεδομένη την παρουσία 8 His σε κάθε μονομερές και την ύπαρξη τριών από αυτές σε διαδοχικές θέσεις στην αλληλουχία της πρωτεΐνης. Η απουσία καταλοίπων Cys από το μόριο μας οδήγησε στη μή επιλογή υδραργυρικών (Hg) ενώσεων στις αρχικές προσπάθειες. Στον πίνακα 13 παρουσιάζονται ενώσεις βαρειών ατόμων που χρησιμοποιήθηκαν στις προσπάθειες για την παραγώγιση των κρυστάλλων της R.*Pvu*II. Για την παραγώγιση ακολουθήθηκαν οι μέθοδοι της συγκρυστάλλωσης και της εμβάπτισης. Αρχική αξιολόγηση της επίδρασης των ενώσεων των βαρειών ατόμων, στην περίπτωση της εμβάπτισης, βασίστηκε σε χρωματικές αλλαγές των κρυστάλλων καθώς και σε ορατές αλλαγές στις επιφάνειες τους. Στην περίπτωση της συγκρυστάλλωσης, εκτός των προηγουμένων, αλλαγές στην φυσιολογική πορεία της κρυστάλλωσης θεωρήθηκαν ενδείξεις για πιθανή ειδική αλληλεπίδραση του βαρέως ατόμου με την πρωτεΐνη. Η επιλογή της συγκέντρωσης της ένωσης του βαρέως ατόμου έγινε έτσι ώστε να είναι η μέγιστη δυνατή χωρίς να προκληθούν ορατές ζημίες (π.χ. ρωγμές) στον κρύσταλλο.

Σε όλες τις περιπτώσεις χρησιμοποιήθηκαν διαλύματα γνωστής συγκέντρωσης των ενώσεων σε νερό ενώ στις περιπτώσεις δυσδιάλυτων ενώσεων χρησιμοποιήθηκαν κεκορεσμένα διαλύματα σε σταθερές συνθήκες και η συγκέντρωση εκφράστηκε ως ποσοστό του κορεσμένου διαλύματος (βλ. κεφάλαιο Δ 4.3.).

Η επιλογή του χρόνου επώασης των κρυστάλλων έγινε επίσης εμπειρικά με βάση τις συνέπειες στην μακροσκοπική ανθεκτικότητα τους σε συνδυασμό με προηγούμενες αναφορές {Blundell & Johnson, 1976} για την χρήση των συγκεκριμένων ενώσεων στην παραγώγιση άλλων πρωτεϊνών. Αξίζει πάντως να σημειωθεί ότι δεν στάθηκε δυνατόν να γίνει προσεκτική μελέτη των συνεπειών του παράγοντα χρόνου καθώς ο χρόνος μεταξύ αρχής της επώασης και των μετρήσεων συνήθως υπαγορευόταν από παράγοντες όπως η διαθεσιμότητα πηγής ακτίνων-Χ, ταξίδια σε πηγές ακτίνων-Χ κλπ. Οι σχετικές τιμές των πινάκων μόνο ενδεικτικά μπορούν να ληφθούν υπόψην.

Πίνακας 13. Ενώσεις βαρειών ατόμων που χρησιμοποιήθηκαν για την δημιουργία παραγώγων των κρυστάλλων της R.*Pvu*II. Στην στήλη αποτέλεσμα δίνονται τιμές ΔF/F για παράγωγα για τα οποία πραγματοποιήθηκαν μετρήσεις σε μέση διακριτικότητα 3-3.5 Å χωρίς να δώσουν χρησιμοποιήσιμες ισόμορφες διαφορές. Με + εκτός των παραγώγων που τελικά χρησιμοποιήθηκαν στον υπολογισμό φάσεων σημειώνονται παράγωγα για τα οποία υπήρξαν ενδείξεις για αλληλεπίδραση των βαρειών ατόμων με την πρωτείνη.

Ενωση	Συγκέντρωση	Χρόνος	Αποτέλεσμα	Μέθοδος
$AgNO_3$	10 mM	2 εβδομάδες	+ Υπ.Φασεων	Συγκρ., Εμβ.
$SmCl_3$	5-10%	2 εβδομάδες	- 12% Ισ.Δ.	Εμβάπτιση
$Sm(NO_3)_3.5H_2O$	10-20 mM (PEG)[2]	1 εβδομάδα	+ αλλίωση κρ.	Εμβάπτιση
$UO_2(AcO)_2$	5-10%	2 εβδομάδες	- 9% Ισ.Δ	Εμβάπτιση
$CsWO_4$	5-10%	2 εβδομάδες	- 13% Ισ.Δ	Εμβάπτιση
$Pr(NO_3)_3.5H_2O$	10-20 mM	1 εβδομάδα	+ Υπ.Φασεων	Εμβάπτιση
Terbium Chloride	10-20 mM	2 μέρες	-	Εμβάπτιση
Holmium oxide	10-20 mM	2 μέρες	-	Εμβάπτιση
$YbCl_3.6H_2O$	10-20 mM	1 εβδομάδα	+ Υπ.Φασεων	Εμβάπτιση
Gold Oxide	10-20 mM	1 εβδομάδα	-	Εμβάπτιση
$KAu(CN)_2$	10-20 mM	1 εβδομάδα	+ αλλίωση κρ.	Εμβάπτιση
Thiomersal	10-20 mM	1 εβδομάδα	+ Υπ.Φασεων	Εμβάπτιση
$Pb(NO_3)_2$	100 mM	2 μέρες	-	Εμβάπτιση
$UO_2(NO_3)_2.H_2O$	10-20 mM (PEG)[2]	1 εβδομάδα	-	Εμβάπτιση
K_2PtCl_4	1-10 mM	2 εβδομάδες	+ 19% Ισ.Δ.	Εμβάπτιση
$Th(NO_3)_4$	10-20 mM (PEG)[2]	2 μέρες	-	Εμβάπτιση

Ισ.Δ. : % Ισόμορφες διαφορές των παράγωγων για τα οποία συλλέχθηκαν δεδομένα.
Υπ. Φάσεων : Χρήση δεδομένων του παραγώγου στον τελικό υπολογισμό φάσεων της πρωτεΐνης.
[2] Μεταφορά των κρυστάλλων για την εμβάπτιση σε διάλυμα PEG8000 10%.

Β 5.3.1. Προκαταρκτικός χαρακτηρισμός των πιθανών παραγώγων.

Στις περισσότερες των περιπτώσεων ο κρυσταλλογραφικός χαρακτηρισμός των κρυστάλλων των πιθανών παραγώγων πραγματοποιήθηκε με την συλλογή δεδομένων σε χαμηλή διακριτικότητα (>=5Å) στο CAD4 περιθλασίμετρο. Ο προσδιορισμός, αρχικά, των παραμέτρων της στοιχειώδους κυψελίδας μας επέτρεψε μια πρώτη εκτίμηση για τον ισομορφισμό των κρυστάλλων. Ακολούθως η συλλογή ενός συνόλου δεδομένων για κάθε πιθανό παράγωγο μας έδωσε την δυνατότητα αφενός να προσδιορίσουμε το ύψος των ισόμορφων διαφορών (βλ. κεφάλαιο Β7.1.), ύστερα από αναγωγή των δεδομένων στην κλίμακα των δεδομένων της φυσικής πρωτεΐνης, και αφετέρου να αναζητήσουμε κορυφές σε χάρτες της συνάρτησης Patterson των διαφορών (βλ. κεφάλαιο Β7.2.).

Η διαδικασία προκαταρκτικού χαρακτηρισμού των κρυστάλλων των παραγώγων που περιγράφηκε πιό πάνω παρουσίασε αρκετές δυσκολίες στην εφαρμογή του. Κυριότερος λόγος γι'αυτό υπήρξε η ελλιπής ποιότητα των δεδομένων που συλλέχθηκαν στο περιθλασίμετρο μιά και το μέγεθος των κρυστάλλων της R.*Pvu*II είναι κατά τεκμήριο μικρό για μετρήσεις σε πηγή χαμηλής ισχύος και χαμηλής ποιότητας δέσμης ακτίνων-Χ. Η διάρκεια των μετρήσεων ακόμα και για ένα σύνολο δεδομένων στα 5.0Å είναι αρκετή (24-36 ώρες) για να προκαλέσει σημαντική φθορά λόγω ακτινοβολίας (radiation damage) στους ευαίσθητους κρυστάλλους των παραγώγων η οποία παρά τις σχετικές διορθώσεις των μετρήσεων αποτελεί σημαντική πηγή σφάλματος. Τέλος, οι υπάρχουσες μεθοδολογίες για την αναγωγή στην ίδια κλίμακα εντάσεων δύο συνόλων δεδομένων παρουσιάζουν σημαντικά σφάλματα όταν τα δεδομένα αυτά προέρχονται από μετρήσεις σε πηγές πολύ διαφορετικής ισχύος (π.χ. CAD4 - σύγχροτρον) και η διαδικασία υπολογισμού των τυπικών αποκλίσεων των εντάσεων είναι διαφορετική {McRee, 1993}.

Η παρουσία των σημαντικών σφαλμάτων που αναφέραμε στις μετρήσεις είχαν ως αποτέλεσμα την πλασματική αύξηση των παρατηρούμενων ισόμορφων διαφορών και την πρόκληση ιδιαίτερου θορύβου στους χάρτες Patterson, που προέκυψαν από αυτές. Προσπάθειες για την αντιμετώπιση του προβλήματος έγιναν στην κατεύθυνση α) της στάθμισης των ανακλάσεων με βάση την τυπική απόκλιση σ και β) με την επιλόγη των ανακλάσεων, στον υπολογισμό των συνθέσεων Patterson, μεγέθους $F>2\sigma(F)$ ή $F>3\sigma(F)$ ανάλογα με την ισχύ του συνόλου δεδομένων, και των διαφορών $\Delta F<4*RMS_{iso}$ ώστε να είναι πιθανές.

Κρύσταλλοι των περισσότερο υποσχόμενων, πιθανών παράγωγων από την προκαταρκτική ανάλυση χρησιμοποιήθηκαν σε μετρήσεις που πραγματοποιήθηκαν σε ισχυρές πηγές ακτίνων-Χ και συγκεκριμένα σε ακτινοβολία σύγχροτρον (Σταθμός ακτινοβολίας X11 του DESY, στο παράρτημα του EMBL στο Αμβούργο) και σε περιστρεφόμενη άνοδο GX21 (30KV, 45mA) στο Ευρωπαικό Εργαστήριο Μοριακής Βιολογίας (EMBL) στην Χαϊδελβέργη σε συνεργασία με τον Dr. Paul Tucker. Στον πίνακα 14 παρουσιάζονται τα στατιστικά των κυριότερων, για την συνέχεια της εργασίας, μετρήσεων παραγώγων. Για το παράγωγο της ένωσης Thiomersal πραγματοποιήθηκε ανεξάρτητη μέτρηση και στις δύο πηγές. Τα δύο σύνολα δεδομένων χρησιμοποιήθηκαν ως διαφορετικά παράγωγα με την ονομασία HG και HG1. Η μέτρηση του HG στην γραμμή ακτινοβολίας X11 έγινε σε μήκος κύματος 0.92Å και τα ζεύγη Bijvoet χρησιμοποιήθηκαν για τον προσδιορισμό των ανώμαλων διαφορών. Η επεξεργασία των δεδομένων τα οποία συλλέχθηκαν με δίσκο ειδώλου από την X11 γραμμή ακτινοβολίας έγινε με το πρόγραμμα Mosflm όπως περιγράφηκε για τα δεδομένα της φυσικής πρωτεΐνης.

Ολα τα παράγωγα για τα οποία πραγματοποιήθηκαν μετρήσεις έδειξαν μη σημαντικές αλλαγές στις διαστάσεις της στοιχειώδους κυψελίδας και η συναρτήσεις Patterson αυτοπεριστροφής, των δεδομένων τους, {δεδομένα που δεν παρουσιάζονται} έδειξαν την ίδια εικόνα με αυτήν της φυσικής πρωτεΐνης (κεφάλαιο Β 6.). Με βάση τα κριτήρια αυτά τα παράγωγα θεωρήθηκαν κατα αρχήν ισόμορφα.

Πίνακας 14. Στατιστικά μετρήσεων σε μέση και ψηλή (<4 Å) διακριτικότητα κρυστάλλων των παραγώγων της πρωτεΐνης με βαριά άτομα.

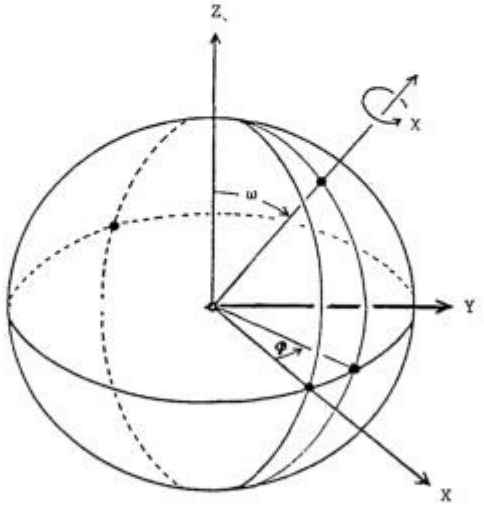
Παράγωγο	Ag	Pr	Yb	Hg	Hg1	Pd	Pt
Ενωση	$AgNO_3$	Pr_2NO_3	$YbSO_4$	Thiomersal	Thiomersal	K_2PdCl_3	$KPtCl_4$
Συγκέντρωση	10mM	20mM	20mM	20mM	20mM	100mM	5mM
Χρόνος	2 Εβδ.	1 Εβδ.	1 Εβδ.	1 Εβδ.	2 Εβδ.	2 Εβδ.	2 Εβδ.
Πηγή	Π.Α.	Π.Α.	Π.Α.	X11	Π.Α.	X11	X11
Ανιχνευτής	Xen.	Xen.	Xen.	MAR	Xen.	MAR	MAR
Μέγιστη Διακριτικότητα (Å)	3.0	2.5	2.5	2.75	2.5	3.0	3.0
a (Å)	84.53	84.61	84.27	83.68	84.27	83.79	84.87
b (Å)	105.68	106.86	106.50	105.68	106.50	105.42	106.48
c (Å)	47.19	47.03	47.06	46.72	47.06	46.92	47.23
R_{sym}	7.4	8.8	6.9	4.0	9.3	6.7	9.0
$I>3\sigma(\%)$	89.32	87.97	87.29	93.8	82.32	84.6	78.1
$I<1\sigma(\%)$	0.50	1.73	1.96	1.9	2.52	5.1	7.0
Αρ. Ανακλάσεων	8620	11127	12666	11020	11480	8659	8514
Πληρότητα%	95.53	72.42	82.44	94.86	74.72	95.96	94.35
Ανωμ. Αρ. Ανακλάσεων	5688	9197	6390	8964	-	7011	6630
Ανωμ. Πληρότητα%	63.03	59.86	41.59	77.16	-	77.70	73.47
$F/\sigma(F)$	14.25	15.69	16.48	33.39	11.53	17.63	13.70

$$Rsymm=\frac{\sum_h (I_h-\overline{I_h})}{\sum \overline{I_h}}$$

Xen : Ανιχνευτής περιοχής Xentronics
MAR : MARearch δίσκος ειδώλου.
Π.Α. : Περιστρεφόμενη Ανοδος
X11 : Γραμμή (beamline) στο DESY/EMBL

Β 6. Προκαταρκτική ανάλυση των δεδομένων της φυσικής πρωτεΐνης: Ο τάξεως-2 άξονας συμμετρίας.

Η παρουσία ενός μοριακού διμερούς ανά ασύμμετρη μονάδα των κρυστάλλων της R.*Pvu*II όπως ήδη δείξαμε καθιστά πιθανή την παρουσία ενός μή κρυσταλλογραφικού άξονα συμμετρίας πιθανά τάξεως-2. Η πληροφορία για την ενδεχόμενη παρουσία του άξονα περιέχεται στα δεδομένα των εντάσεων τόσο της φυσικής πρωτεΐνης όσο και των παραγώγων με βαριά άτομα. Αναζητήσαμε λοιπόν τον άξονα αυτόν με την χρήση της συνάρτησης αυτοπεριστροφής Patterson (Self Rotation Function) {Blow, 1985}, {Lattman, 1985}. Η ερμηνεία των αποτελεσμάτων της συνάρτησης αυτής έχει ως στόχο των προσδιορισμό της διεύθυνσης του άξονα μέσα στην ασύμμετρη μονάδα. Στο πολικό σύστημα τρείς γωνίες φ,ω (ψ) και κ (χ) είναι απαραίτητες για τον ορισμό του άξονα, με τις δυό πρώτες να καθορίζουν την διεύθυνση του άξονα και την γωνία κ να προσδιορίζει την περιστροφή γύρω από αυτόν προκειμένου το ένα μονομερές να συμπέσει με το δεύτερο (εικόνα 13).



Εικόνα 13. Παράσταση του ορισμού των φ, ω (ψ) και κ (χ) γωνιών στο πολικό σύστημα .

Οι όροι της χαμηλής διακριτικότητας αποκλείστηκαν από τον υπολογισμό της ολοκλήρωσης της συνάρτησης αυτοπεριστροφής καθώς κυρίως περιγράφουν της περιοχές του διαλύτη. Η συνάρτηση αυτοπεριστροφής Patterson πραγματοποιήθηκε με το πρόγραμμα POLARRFN {CCP4, 1979}.

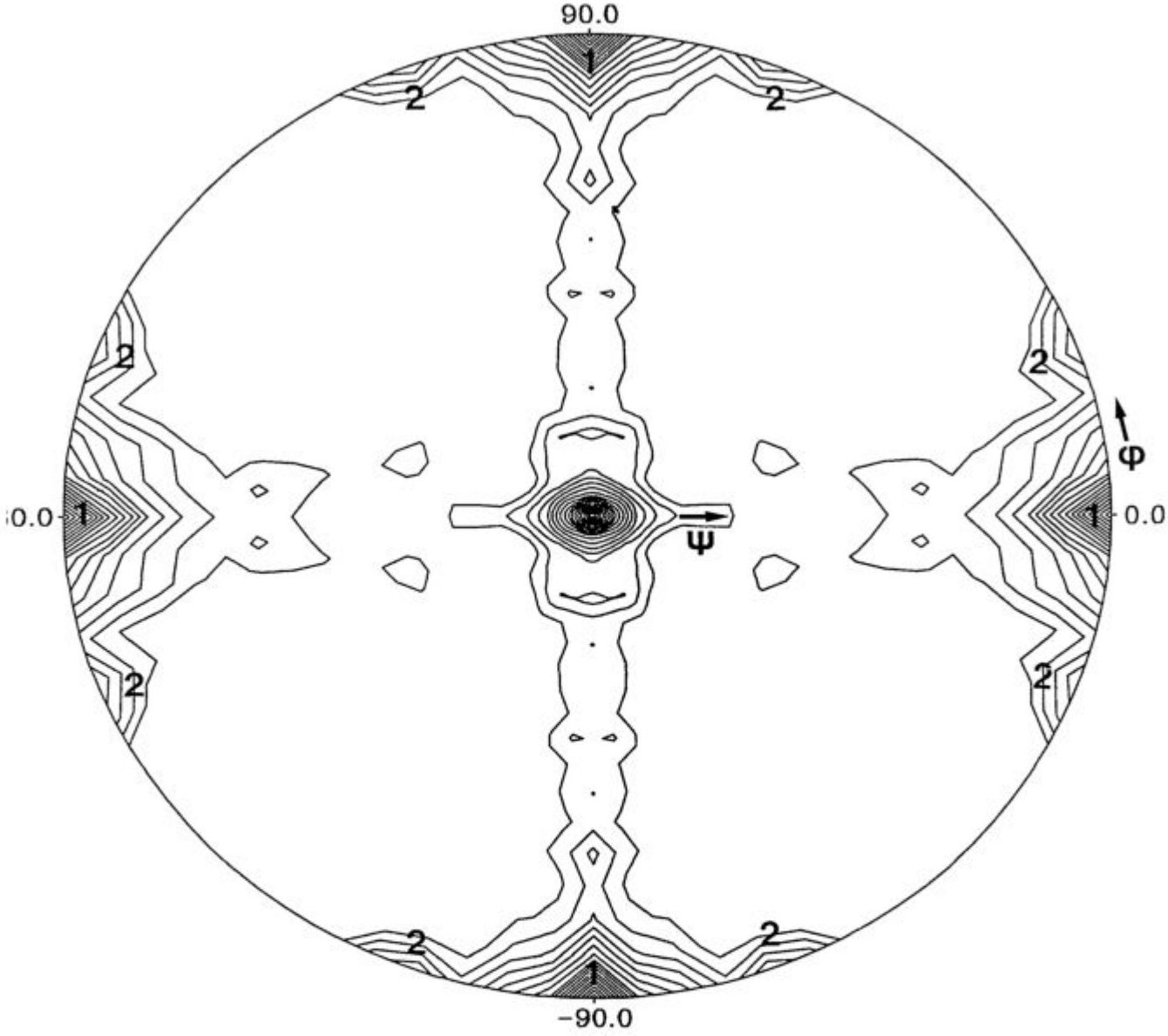
Η συνάρτηση Patterson αποτελεί ένα σύνολο ανυσμάτων μεταξύ των ατόμων του μορίου. Τα ανύσματα αυτά μπορεί να αντιπροσωπεύουν

ενδομοριακές αποστάσεις αλλά και αποστάσεις μεταξύ ατόμων διαφορετικών μορίων. Η επιλογή της ακτίνας ολοκλήρωσης της συνάρτησης αυτοπεριστροφής Patterson γίνεται έτσι ώστε να ελαχιστοποιείται η συμμετοχή ανυσμάτων μεταξύ ατόμων διαφορετικών μορίων (cross vectors). Στην πράξη κάνοντας την υπόθεση ότι η υπό μελέτη πρωτεΐνη είναι ένα περίπου σφαιρικό μόριο και επιλέγοντας ως ακτίνα ολοκλήρωσης μία τιμή ακτίνας 75-80% του μεγέθους του μορίου {Blow, 1985} επιτυγχάνουμε τον αποκλεισμό των περισσότερων διαμοριακών ανυσμάτων. Στον πίνακα που ακολουθεί συνοψίζονται τα αποτελέσματα για δύο διαφορετικές ακτίνες ολοκλήρωσης της συνάρτησης ενώ στην εικόνα 14 παρουσιάζεται η τομή του χάρτη της συνάρτησης για γωνία $\kappa=180^{o}$.

Πίνακας 15. Κορυφές της συνάρτησης αυτοπεριστροφής Patterson. Η επιλογή του κάτω ορίου της διακριτικότητας έγινε ανάλογα με την ακτίνα ολοκλήρωσης, και με βάση των εμπειρικό τύπο 2*π*ακτίνα ολοκλήρωσης/max. διακριτικότητας <37.67 για λόγους ορίων του προγράμματος σε αριθμό ανυσμάτων. Το ύψος των κορυφών παρουσιάζεται ως ποσοστό της κορυφής από την σύμπτωση του μορίου με τον εαυτό του (origin). Σε όλες τις περιπτώσεις η κορυφή που αντιστοιχεί σε άξονα με $\omega=90^{o}$ και $\varphi\sim20^{o}$ εμφανίστηκε πρώτη σε ύψος. Για λόγους σύγκρισης αναφέρεται το ύψος της δεύτερης κορυφής σε κάθε περίπτωση.

Ακτίνα ολοκλήρωσης (Å)	Διακριτικότητα (Å)	ω	φ	κ	Σχετικό ύψος (%)	Δεύτερη κορυφή (%)
20Å	8-3.5	90°	19.4^{o}	180°	38.5	20.1
30Å	8-5.16	90°	17.8^{o}	180°	34.7	26.5

Παρόμοια υπήρξαν τα αποτελέσματα για σύνολα δεδομένων από κρυστάλλους παραγώγων με βαριά άτομα. Συνοψίζοντας, τα παραπάνω αποτελέσματα υποδεικνύουν κατά σαφή τρόπο την ύπαρξη ενός τάξεως-2 άξονα παράλληλου με το επίπεδο a, b και με κλίση 20° περίπου, από τον άξονα a. Ο προσδιορισμός της διεύθυνσης του τάξεως-2 άξονα με μεγαλύτερη ακρίβεια καθώς και η μετατόπιση του σε σχέση με την αρχή των αξόνων όπως θα δούμε στη συνέχεια έγινε με βάση τις θέσεις των βαρειών ατόμων στα παράγωγα.



Εικόνα 14. Συνάρτηση αυτοπεριστροφής Patterson σε όρια 8-3.5Å. *Στερεογραφική προβολή περιστροφών με κ=180°. Διακρίνονται οι κορυφές που αντιστοιχούν σε φ=19.4^{o} και ω (ψ) =90°. Ισουψείς ανά 5 μονάδες στην περιοχή15-100 όπου το 100 αντιστοιχεί στη σύμπτωση του μορίου με τον εαυτό του (origin). Οί κορυφές με την ένδειξη 1 αντιπροσωπεύουν κορυφές αρχής (origin peaks) και οι κορυφές 2 αντιστοιχούν στον άξονα της μη κρυσταλλογραφικής συμμετρίας των κρυστάλλων της R.PvuII Η κατεύθυνση αυξήσης των φ και ψ (ω) γωνιών στο διάγραμμα σημειώνεται με βέλη.*

Β 7. Πολλαπλή ισόμορφη αντικατάσταση. (Multiple Isomorphous Replacement (MIR) phasing)

Β 7.1 Αναγωγή δεδομένων των παραγώγων με βαριά άτομα στην κλίμακα των δεδομένων των κρυστάλλων της φυσικής πρωτεΐνης. Στατιστική ανάλυση των διαφορών των δεδομένων παραγώγων με βαριά άτομα - φυσικής πρωτεΐνης.

Την συλλογή των δεδομένων των παραγώγων της πρωτεΐνης με βαριά άτομα ακολούθησε αναγωγή των δεδομένων στην κλίμακα των δεδομένων της φυσικής πρωτεΐνης. Η μέθοδος που χρησιμοποιήθηκε είναι αυτή της ανισότροπης αναγωγής όπου ο συντελεστής αναγωγής για μία ανάκλαση [hkl] δίνεται από τον τύπο (...1). Στον τύπο αυτό C ο συνολικός παράγοντας αναγωγής και B** οι έξι ανισοτροπικοί παράγοντες αναγωγής. Την ανισότροπη αναγωγή ακολούθησε η εφαρμογή αναγωγής Wilson με συντελεστή $K= (\Sigma F_P^2/\Sigma F_{PH}^2)^{1/2}$ επιτυγχάνοντας $<F_{PH}^2>=<F_P^2>$ όπου F_P το πλάτος των συντελεστών δομής της φυσικής πρωτεΐνης και F_{PH} του παραγώγου. Η αναγωγή Wilson έχει επίδραση μόνο πάνω στον συνολικό παράγοντα αναγωγής C ενώ αφήνει αναλλοίωτους τους παράγοντες B από την ανισότροπη αναγωγή.

$C*\exp(-(h^2*B11+k^2*B22+l^2*B33+2hk*B12+2hl*B13 +2kl*B23))$ (...1) {CCP4, 1979}

(B_{12}, B_{13}, B_{23}=0 για ορθορομβικές ομάδες χώρου.)

Η εφαρμογή των προηγουμένων πραγματοποιήθηκε με το πρόγραμμα SCALEIT του CCP4 πακέτου προγραμμάτων για όλα τα παράγωγα και έγινε με βάση το NatM σύνολο δεδομένων της φυσικής πρωτεΐνης (πίνακας 12). Η ίδια αναγωγή εφαρμόστηκε στις περιπτώσεις όπου ήταν διαθέσιμα δεδομένα ανώμαλου σκεδασμού. Στον πίνακα 16 παρουσιάζονται τα στατιστικά της αναγωγής για όλα τα παράγωγα για τα οποία η διακριτικότητα του συνόλου δεδομένων ξεπερνάει τα 4Å. Ο παράγοντας $R_{iso}=\Delta F/F=(<(F_{PH}-F_P)/F_P>)$ αποτελεί δείκτη του μεγέθους των διαφορών φυσικής πρωτεΐνης - παραγώγου. Τα παράγωγα του Pd και Pt παρά το γεγονός ότι έδειξαν αξιόλογες ισόμορφες διαφορές κατά την διαδικασία της αναγωγής δεν αποδείχθηκαν χρήσιμα για τον προσδιορισμό των φάσεων της πρωτεΐνης. Οι συνθέσεις Patterson των ισόμορφων διαφορών για τα παράγωγα αυτά υπήρξαν μή ερμηνεύσιμες πιθανά λόγω της ύπαρξης πολλαπλών θέσεων πρόσδεσης (με χαμηλή πληρότητα (occupancy) για την Pt). Από τις τιμές του K_{emp} του πίνακα 16 είναι φανερό το σημαντικό μέγεθος των ανώμαλων διαφορών του

Pr που αποδείχθηκαν ιδιαίτερα χρήσιμες στην συνέχεια. Τα ανώμαλα δεδομένα του Hg αν και ιδιαίτερα ασθενή ήταν αρκετά ώστε να δώσουν ερμηνεύσιμη σύνθεση Patterson και να επιβεβαιώσουν τα αποτελέσματα των ισόμορφων διαφορών. Ελάχιστη εκτίμηση του αριθμού των προσδεδεμένων βαρειών ατόμων στην πρωτεΐνη μπορεί να δοθεί από τον τύπο (...2),

$\sqrt{N_h} = (\sqrt{MW}) * R_{iso} / (0.39 * f_h)$ (...2) {McRee, 1993}

με τον τύπο αυτό και τις τιμές $R_{iso}=\Delta F/F$ του πίνακα 16 περιμένουμε πλήρως κατειλημμένες 4 θέσεις για τα Yb και Pr, 2 για τον Hg και 5 για τον Ag τιμές που είναι σε καλή συμφωνία με την μετέπειτα ανάλυση. Για το Pd περιμένουμε το ελάχιστο 7 θέσεις πρόσδεσης γεγονός που πιθανά εξηγεί την πολυπλοκότητα που εμφάνισε η σύνθεση Patterson των ισόμορφων διαφορών του. Στήν εικόνα 15 παρουσιάζεται διάγραμμα ΔF/F για το σύνολο των παραγώγων που χρησιμοποιήθηκαν στον προσδιορισμό των φάσεων, η κορυφή σε διακριτικότητα περίπου 6.2Å οφείλεται σε πτώση των εντάσεων της φυσικής πρωτεΐνης {Leslie, 1991} ενώ η σταδιακή αύξηση του λόγου με την αύξηση της διακριτικότητας σχετίζεται με την αύξηση του σφάλματος των μετρήσεων σε συνδυασμό με την πτώση της έντασης των ανακλάσεων.

Πίνακας 16. Στατιστικά της ανισότροπης αναγωγής των δεδομένων των παραγώγων της πρωτεΐνης με βαριά άτομα. Η αναγωγή πραγματοποιήθηκε στα όρια διακριτικότητας 20-3Å για όλα τα παράγωγα.

Παράγωγο	Ag	Pr	Yb	Hg	Hg1	Pd	Pt
Μέγιστη Διακριτικότητα (Å)	3.0	2.5	2.5	2.7	2.5	3.0	3.0
R_{iso}[1]	22.5	24.2	30.0	22.5	25.6	23.9	18.4
$rms\Delta_{ισο}$*	815	788	957	690	797	865	649
$rms\Delta_{ανωμ}$**	343	348	246	138	-	255	334
K_{emp}[2]	4.70	3.04	6.14	8.92	-	6.22	3.87
N_h[3]	5	4	4	2	2	7	1

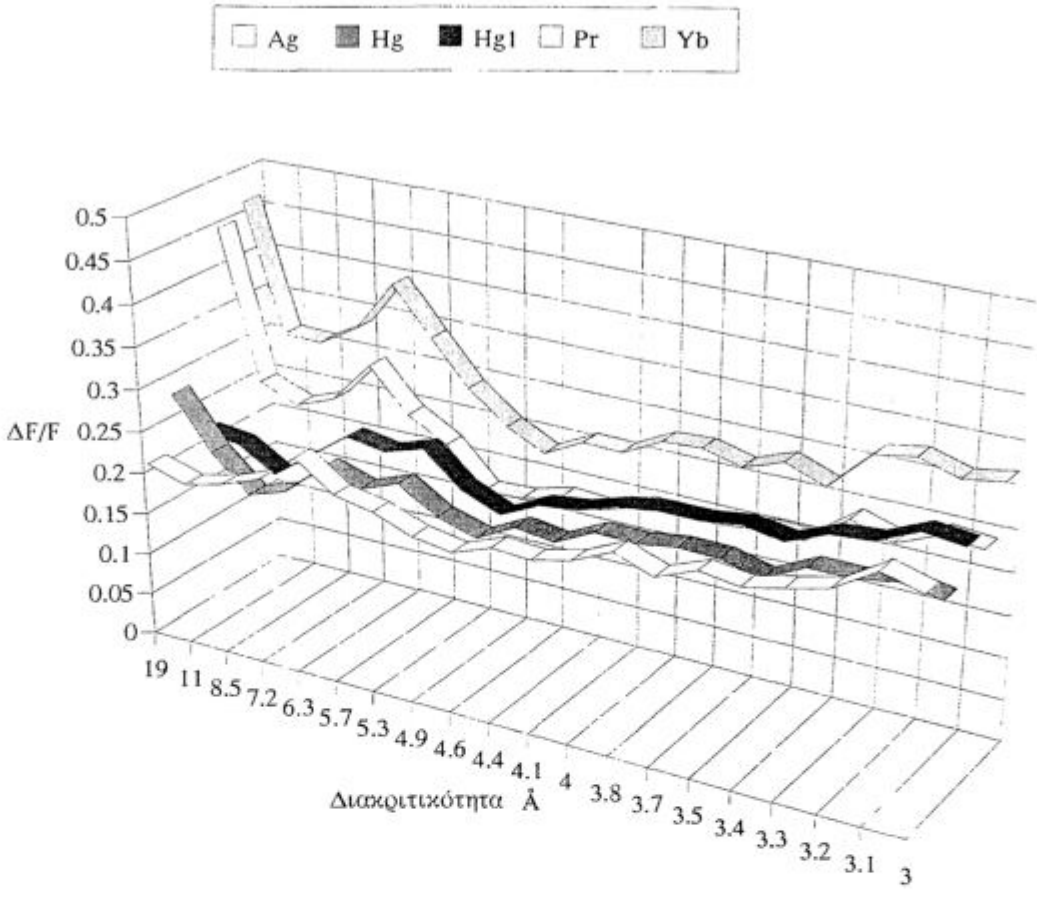
[1] $R_{iso} = \Delta F/F = \langle | F_{PH} - F_P | \rangle / \langle | F_P | \rangle$

[2] $K_{emp} = 2\sqrt{\left(\frac{\langle (F_{PH}-F_P)^2 \rangle}{\langle (F_{PH}(+)-F_{PH}(-))^2 \rangle}\right)}$

[3] Αναμενόμενος αριθμός πλήρως κατειλημμένων θέσεων πρόσδεσης.

* $rms\Delta_{ισο}$: $rms(F_{PH}-F_P)$

** $rms\Delta_{ανωμ}$: $rms(F_{PH}(+)-F_{PH}(-))$



Εικόνα 15. Διακύμανση ΔF/F αυξανομένης της διακριτικότητας για τα παράγωγα Ag, Pr, Hg, Hg1, Yb (πίνακας 14).

Β 7.2. Προσδιορισμός και βελτιστοποίηση των θέσεων βαρειών ατόμων.

Η επίλυση του προβλήματος προσδιορισμού των φάσεων στην MIR μέθοδο {Watenpaugh, 1985} απαιτεί τον καθορισμό των θέσεων πρόσδεσης των βαρειών ατόμων, για δύο τουλάχιστον διαφορετικά παράγωγα των κρυστάλλων της φυσικής πρωτεΐνης.

Προκειμένου να προσδιορίσουμε τις συντεταγμένες των θέσεων πρόσδεσης των βαρειών ατόμων στην ασύμμετρη μονάδα των κρυστάλλων των παραγώγων προχωρήσαμε στον υπολογισμό των συνθέσεων Patterson {Patterson, 1934} ισόμορφων διαφορών με συντελεστές $(|F_{PH}|-|F_P|)^2$ {Perutz, 1956}, {Blow, 1958} σε διακριτικότητα 20Å- 4Å η 5Å. Παράλληλα, όπου δεδομένα ανώμαλου σκεδασμού ήταν διαθέσιμα, πραγματοποιήθηκαν συνθέσεις Patterson με την χρήση των διαφορών της ανώμαλης σκέδασης με συντελεστές δηλαδή $(|F_{PH}|_+- |F_{PH}|_-)^2$ σε διακριτικότητα 3Å.

Ανάλυση των τομών Harker {Harker, 1936} των χαρτών Patterson μας επέτρεψε των προσδιορισμό των κύριων θέσεων πρόσδεσης για κάθε παράγωγο. Για την χωροομάδα $P2_12_12$ ορίζονται τρείς Harker τομές : α) u=1/2, β) v=1/2 και γ) w=0. Η u=1/2 τομή περιέχει κορυφές των οποίων οι συντεταγμένες συνδέονται με αυτές του βαρέως ατόμου ως εξής v=1/2±2y και w=±2z και αντίστοιχα για τις άλλες δύο τομές u=1/2±2x, w=±2z και u=±2x, v=±2y. Στην χωροομάδα $P2_12_12$ κάθε τριπλέτα κορυφών από τις Harker τομές, που αντιστοιχεί σε μία θέση πρόσδεσης, μπορεί να ερμηνευτεί με 8 διαφορετικούς τρόπους οι οποίοι ισοδυναμούν με διαφορετική επιλογή της αρχής των αξόνων (origin). Για τον λόγο αυτό πέρα από την ανάλυση των τομών harker για τον προσδιορισμό των μεμονωμένων θέσεων, ανύσματα μεταξύ των υποψηφίων θέσεων (cross vectors) χρησιμοποιήθηκαν για την επιλογή των συντεταγμένων των θέσεων των βαρειών ατόμων σε αναφορά με το ίδιο σημείο αρχής των αξόνων. Στις εικόνες 17-25 παρουσιάζονται οι τομές harker από τις συνθέσεις Patterson των ισόμορφων και ανώμαλων διαφορών για τα 5 σύνολα δεδομένων των παράγωγων που χρησιμοποιήθηκαν για την επίλυση του προβλήματος των φάσεων της R.*Pvu*II. Στην διαδικασία ερμηνείας των χαρτών της Patterson σύνθεσης των παραγώγων χρησιμοποιήθηκε το πρόγραμμα αυτόματης ερμηνίας HASSP {CCP4, 1979}, επίλυση με το "χέρι" ύστερα από καταγραφή των κορυφών της τρισδιάστατης σύνθεσης με το προγραμμα PEAKMAX {CCP4, 1979} και τέλος λογισμικό {Athanasiadis, πέρα από το αντικείμενο της παρούσης διατριβής} βασισμένο στην μέθοδο ψαξίματος πλέγματος (grid searching) {Jones & Stuart, 1991}.

Οι θέσεις αυτές βελτιστοποιήθηκαν ανεξάρτητα για κάθε παράγωγο με την χρήση των κεντροσυμμετρικών ανακλάσεων μόνο με το πρόγραμμα REFINE2

{CCP4, 1979}. Στην βελτιστοποίηση με βάση τις κεντροσυμμετρικές ανακλάσεις επιδιώκεται η ελαχιστοποίηση της ποσότητας $E=\Sigma[(F_{H(υπ)}-F_{H(παρ)}]^2$ όπου $F_{H(παρ)}= |F_{PH}-F_P|$ {Sparks, 1985}. Ακολούθησε προσδιορισμός των φάσεων της πρωτεΐνης με την χρήση των ισόμορφων και ανώμαλων δεδομένων του παραγώγου Pr το οποίο έδωσε τα καλύτερα στατιστικά στην πρώτη φάση της βελτιστοποίησης. Οι φάσεις αυτές χρησιμοποιήθηκαν για (cross) Fourier συνθέσεις των διαφορών (συντελεστές $m*(F_{PH}-F_P)*\exp(i\alpha_P)$) των υπολοίπων παραγώγων προκειμένου να επιλεχθούν οι θέσεις των διαφορετικών παραγώγων σε αναφορά με το ίδιο σχετικό σημείο αρχής των αξόνων (από τις 8 επιλογές για την $P2_12_12$ χωροομάδα) και την επίλυση του προβλήματος της επιλογής του ίδιου εναντιομερούς για κάθε παράγωγο {Mathews, 1966}. Η τελική επιλογή του εναντιομερούς για το σύνολο των παραγώγων πραγματοποιήθηκε μετά τον προσδιορισμό των φάσεων και τον υπολογισμό χαρτών ηλεκτρονιακής πυκνότητας της πρωτεΐνης με βάση την εμφάνιση στον χάρτη δεξιόστροφων α-ελίκων και αμινοξέων της L διαμόρφωσης. Στην συνέχεια οι νέες θέσεις πρόσδεσης των βαρειών ατόμων χρησιμοποιήθηκαν για βελτιστοποίηση των παραμέτρων των βαρειών ατόμων με την μέθοδο "maximum likelihood phase" με την χρήση του προγράμματος MLPHARE {Otwinowski, 1991}. Στην "maximum likelihood phase" μέθοδο υπολογίζονται συντελεστές για όλες τις δυνατές φάσεις φ_P μιάς ανάκλασης και ελαχιστοποιούνται οι σταθμισμένες διαφορές μεταξύ των πειραματικών και υπολογισμένων τιμών του F_{PH} (με τις φάσεις αυτές), λαμβάνοντας υπόψην τις παραμέτρους των βαρειών ατόμων. Η μέθοδος αυτή βελτιστοποιεί ταυτόχρονα όλες τις παραμέτρους των βαρειών ατόμων και έχει το πλεονέκτημα του μή επηρεασμού των φάσεων που προκύπτουν από πιθανές μεμονωμένες λανθασμένες θέσεις πρόσδεσης.

Με τις φάσεις της πρωτεΐνης που προέκυψαν έγιναν συνθέσεις Fourier για κάθε παράγωγο με τις υπολειπόμενες διαφορές $(F_{PH}-F_P)-F_C$ για τον εντοπισμό δευτερευόντων θέσεων πρόσδεσης βαρειών ατόμων. Οι παράμετροι των θέσεων πρόσδεσης βαρειών ατόμων όπως προέκυψαν απο την συνολική διαδικασία βελτιστοποίησης παρουσιάζονται στον πίνακα 17 και τα στατιστικά του προσδιορισμού των φάσεων στον τελευταίο κύκλο της βελτιστοποίησης δίνονται στον πίνακα 18.

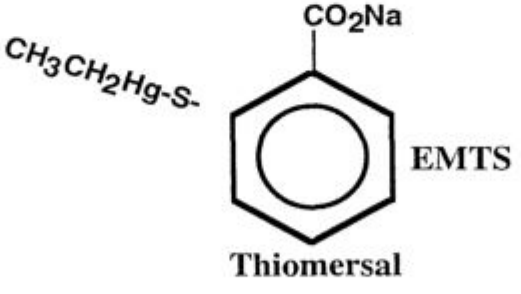
Πρασεοδύμιο-Υττέρβιο.

Για το παράγωγο του Pr, από την ανάλυση των Harker τομών της συνάρτησης Patterson διαφορών προσδιορίστηκαν δύο θέσεις πρόσδεσης (θέσεις 1-2, εικόνα 17). Σε πλήρη συμφωνία ήταν τα αποτελέσματα από την ανάλυση των Harker τομών της Patterson των ανώμαλων διαφορών (εικόνα 18). Η παρουσία και των τεσσάρων ανυσμάτων μεταξύ των βαρειών ατόμων μας επέτρεψε την

επιλογή του σωστού συνδυασμού των συντεταγμένων για τις δύο θέσεις. Στο σημείο αυτό έγινε και η επιλογή της αρχής των αξόνων καθώς το παράγωγο του Pr χρησιμοποιήθηκε στην συνέχεια ως σημείο αναφοράς για την αντίστοιχη επιλογή για τα υπόλοιπα παράγωγα. Την βελτιστοποίηση των θέσεων αυτών ακολούθησε σύνθεση Fourier διαφορών από την οποία προέκυψαν οι δευτερεύουσες θέσεις 3-4, γειτονικές προς της κύριες θέσεις και των οποίων η παρουσία επιβεβαιώθηκε από χάρτες Patterson στα 3Å. Οι τέσσερεις θέσεις πρόσδεσης του Pr βρέθηκε νά συνδέονται ανά δύο με άξονα ίδιας διεύθυνσης με αυτόν που προβλεπόνταν από την συνάρτηση αυτοπεριστροφής. Σχεδόν πανομοιότυπους χάρτες Patterson πήραμε για το παράγωγο του υττερβίου (εικόνα 19) γεγονός που αποδίδεται στις παραπλήσιες χημικές ιδιότητες των δύο μετάλλων, καθώς και τα δύο ανήκουν στην ομάδα των Λανθανίδων. Λιγότερο σαφής υπήρξε ο χάρτης της συνάρτησης Patterson των ανώμαλων διαφορών του υττερβίου (εικόνα 20) πράγμα το οποίο θα πρέπει να αποδοθεί κυρίως στο γεγονός της μή πληρότητας του συνόλου των ανώμαλων δεδομένων (41.59% πληρότητα μέχρι τα 2.5 Å διακριτικότητα).

Thiomersal.

Δύο ανεξάρτητα σύνολα δεδομένων συλλέχθηκαν για το υδραργυρικό αυτό παράγωγο (εικόνα 16). Δεδομένα ανώμαλου σκεδασμού συλλέχθηκαν στην μέτρηση που πραγματοποιήθηκε στήν γραμμή ακτινοβολίας X11 του DESY. Δύο θέσεις πρόσδεσης προσδιορίστηκαν από τις συνθέσεις Patterson των ισόμορφων διαφορών (εικόνες 21,23) ενώ η σαφέστερη εικόνα δόθηκε από την σύνθεση Patterson των ανώμαλων διαφορών (εικόνα 22). Με συνθέσεις Fourier διαφορών δεν στάθηκε δυνατό να εντοπιστούν σημαντικές δευτερεύουσες θέσεις πρόσδεσης.



Εικόνα 16. Συντακτικός τύπος της υδραργυρικής ένωσης Thiomersal (αιθυλο-υδραργυρο-θείο-σαλικυλικό οξύ, EMTS). *(Blundell & Johnson, 1976)*

Νιτρικός Αργυρος.

Ιδιαίτερης πολυπλοκότητας παρουσιάστηκαν οι συνθέσεις της συνάρτησης Patterson για το παράγωγο του αργύρου (εικόνες 24,25) που υπήρξε το πρώτο από

τα χρήσιμα παράγωγα των κρυστάλλων της R.*Pvu*II πρωτεΐνης. Η παρουσία πολλαπλών θέσεων πρόσδεσης είχε ως αποτέλεσμα την ελάττωση του λόγου σήματος προς θόρυβο της συνθέσεως Patterson. Οι θέσεις 1 και 2 προσδιορίστηκαν από την Patterson ισόμορφων διαφορών ενώ άλλες 7 θέσεις πρόσδεσης επιλέχθηκαν από συνθέσεις Fourier διαφορών και υπολοιπόμενων διαφορών με φάσεις που προέκυψαν από τα υπόλοιπα παράγωγα. Ολες οι δευτερεύουσες θέσεις επιλέχθηκαν χρησιμοποιώντας ως επιπλέον κριτίριο την ικανοποίηση του μη κρυσταλλογραφικού άξονα διπλής συμμετρίας όπως αυτός προσδιορίστηκε με ακρίβεια από τις θέσεις των υπολοίπων παραγώγων.

Πίνακας 17. Παράμετροι των βαρειών ατόμων για τα παράγωγα που χρησιμοποιήθηκαν στον υπολογισμό των φάσεων της R.*Pvu*II όπως προέκυψαν ύστερα από βελτιστοποίηση με το πρόγραμμα MLPHARE. Οι συντεταγμένες των θέσεων δίνονται σε κλασματικές τιμές. Στην στήλη Συμ. αναφέρεται το συμμετρικό με βάση την μη-κρυσταλλογραφική συμμετρία άτομο για κάθε θέση ενώ με = συμβολίζονται δευτερεύουσες θέσεις γειτονικές προς άλλες κύριες θέσεις πρόσδεσης.

Παράγ.	Θέση	x	y	z	Σχετική πληρότητα	B (Å^2)	Αν. Πληρ.	Συμ.
Ag	1	0.172	0.239	0.202	9.511	80.35	8.31	-
Ag	2	0.014	0.485	0.179	5.994	79.00	4.57	->4
Ag	3	0.482	0.399	0.455	5.523	91.33	5.12	->6
Ag	4	0.086	0.615	0.575	3.442	44.25	3.44	->2
Ag	5	0.375	0.428	0.024	4.086	72.45	2.93	->7
Ag	6	0.491	0.407	0.324	3.351	155.47	4.24	->3
Ag	7	0.383	0.457	0.741	2.156	47.35	1.38	->5
Ag	8	0.004	0.487	0.134	2.077	60.91	1.84	=2
Ag	9	0.422	0.118	0.370	2.493	51.00	1.66	=4
Hg	1	0.293	0.552	0.497	5.395	40.24	4.39	->2
Hg	2	0.220	0.409	0.261	4.823	42.45	3.72	->1
Hg1	1	0.292	0.552	0.498	5.867	40.72	-	->2
Hg1	2	0.220	0.409	0.261	4.773	35.31	-	->1
Pr	1	0.205	0.569	0.702	9.963	35.68	8.41	->2
Pr	2	0.146	0.447	0.061	11.254	37.32	9.14	->1
Pr	3	0.112	0.453	0.079	3.417	18.28	0.73	->4
Pr	4	0.179	0.578	0.694	3.748	48.98	1.86	->3
Yb	1	0.207	0.570	0.702	9.490	31.73	8.758	->2
Yb	2	0.146	0.447	0.062	12.227	38.09	9.363	->1
Yb	3	0.113	0.452	0.078	3.465	9.38	1.042	->4
Yb	4	0.184	0.576	0.697	3.555	34.27	2.052	->3

Πίνακας 18. Στατιστικά του τελευταίου κύκλου βελτιστοποίησης των παραμέτρων των βαρειών ατόμων με το πρόγραμμα MLPHARE.

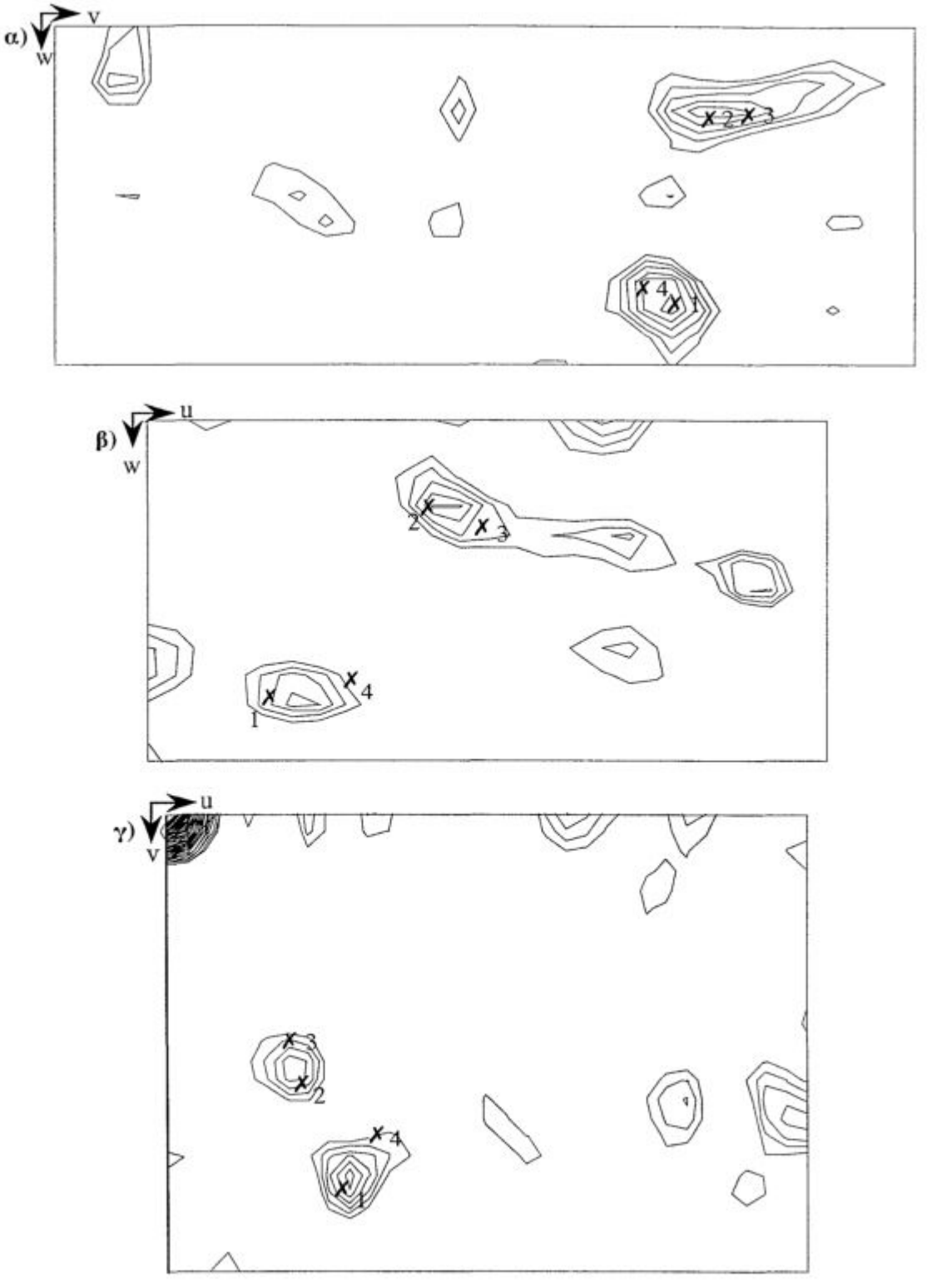
Παράγ.	Ορια Διακρ. (Å)	Αρ. Ανακλ.	Cullis R [1]	Cullis R An.[2]	Rms Ισχύς Πρ. Φάσεων[3] (acen)§	Rms E (acen)§
Ag	∞-3	8464	0.86	1.01	0.55	668
Hg	∞-2.8	10438	0.84	1.11	0.74	553
Pr	∞-2.8	9603	0.61	1.16	1.62	454
HgI	∞-2.8	9253	0.84	-	0.74	619
Yb	∞-2.8	10146	0.71	1.06	1.47	603

[1] R_{Cullis}=<Σφαλμα ανυσματικού αθροίσματος>/<ΔF>

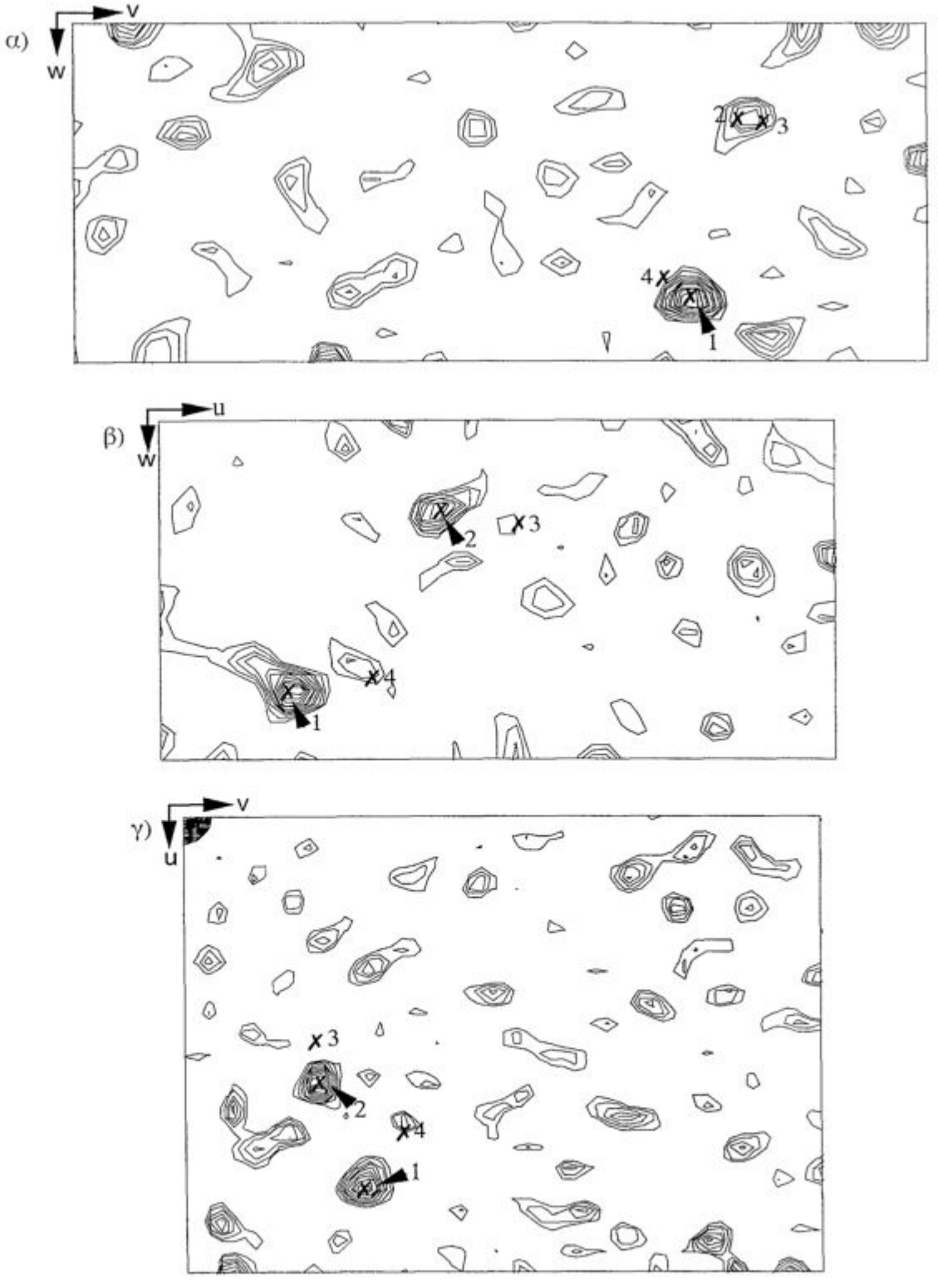
[2] Ανώμαλα δεδομένα: R_{Cullis}=<Δανωμ>/<Σφάλμα ανυσματικού αθροίσματος για αν. δεδομένα>

[3] Rms Ισχύς Προσδιορισμού Φάσεων= $RmsF_H$/Σφάλμα ανυσματικού αθροίσματος (E)

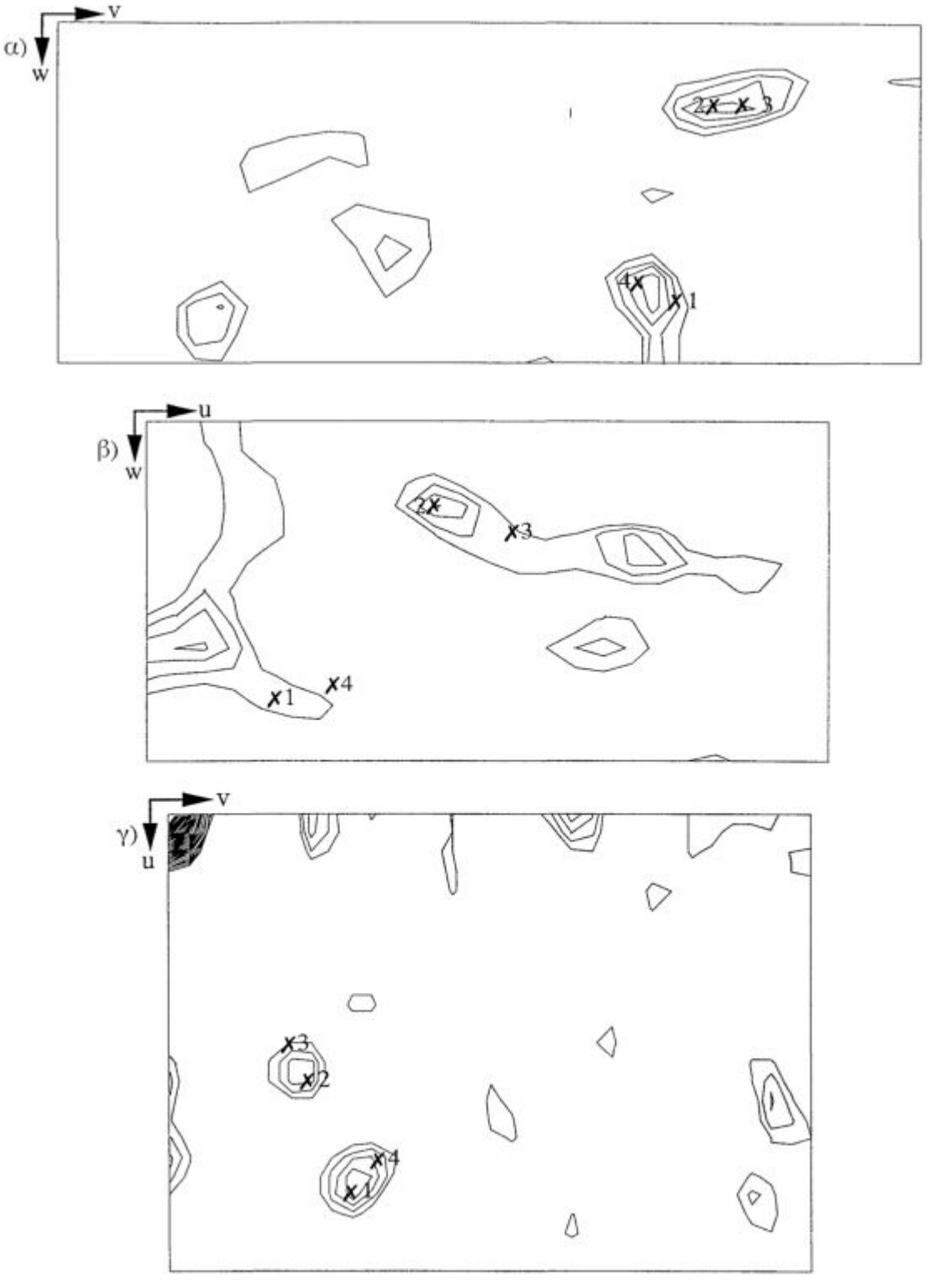
§ μη κεντροσυμμετρικές ανακλάσεις.



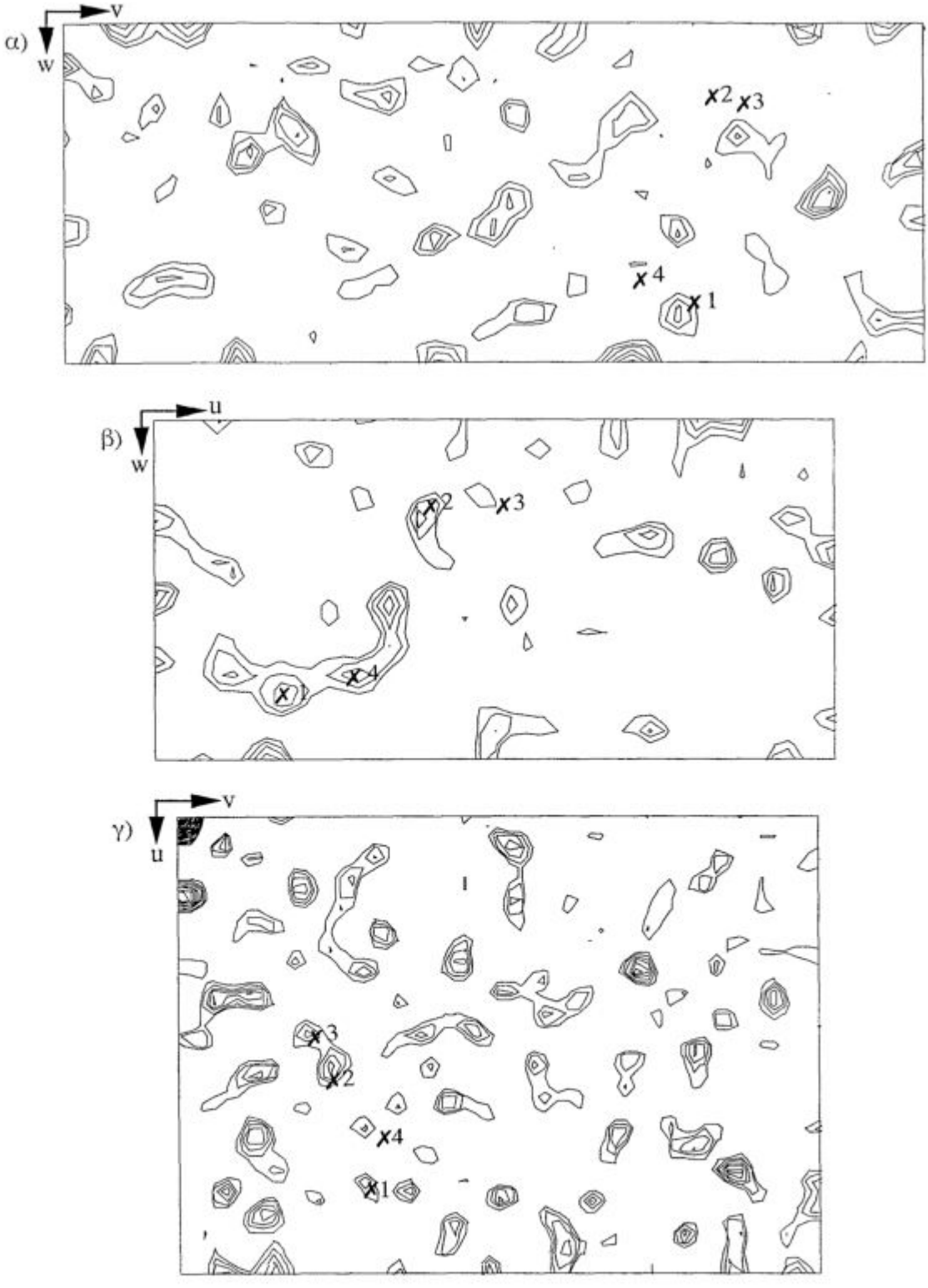
Εικόνα 17. *20-5Å Συνάρτηση διαφορών Patterson για το παράγωγο του Προασεοδυμίου. α) u=1/2 β) v=1/2 γ) w=0. Ισουψείς απο το 1σ ανά 1/2σ. 1-2 οι κύριες θέσεις πρόσδεσης και 3-4 οι δευτερέυουσες (πίνακας 17).*



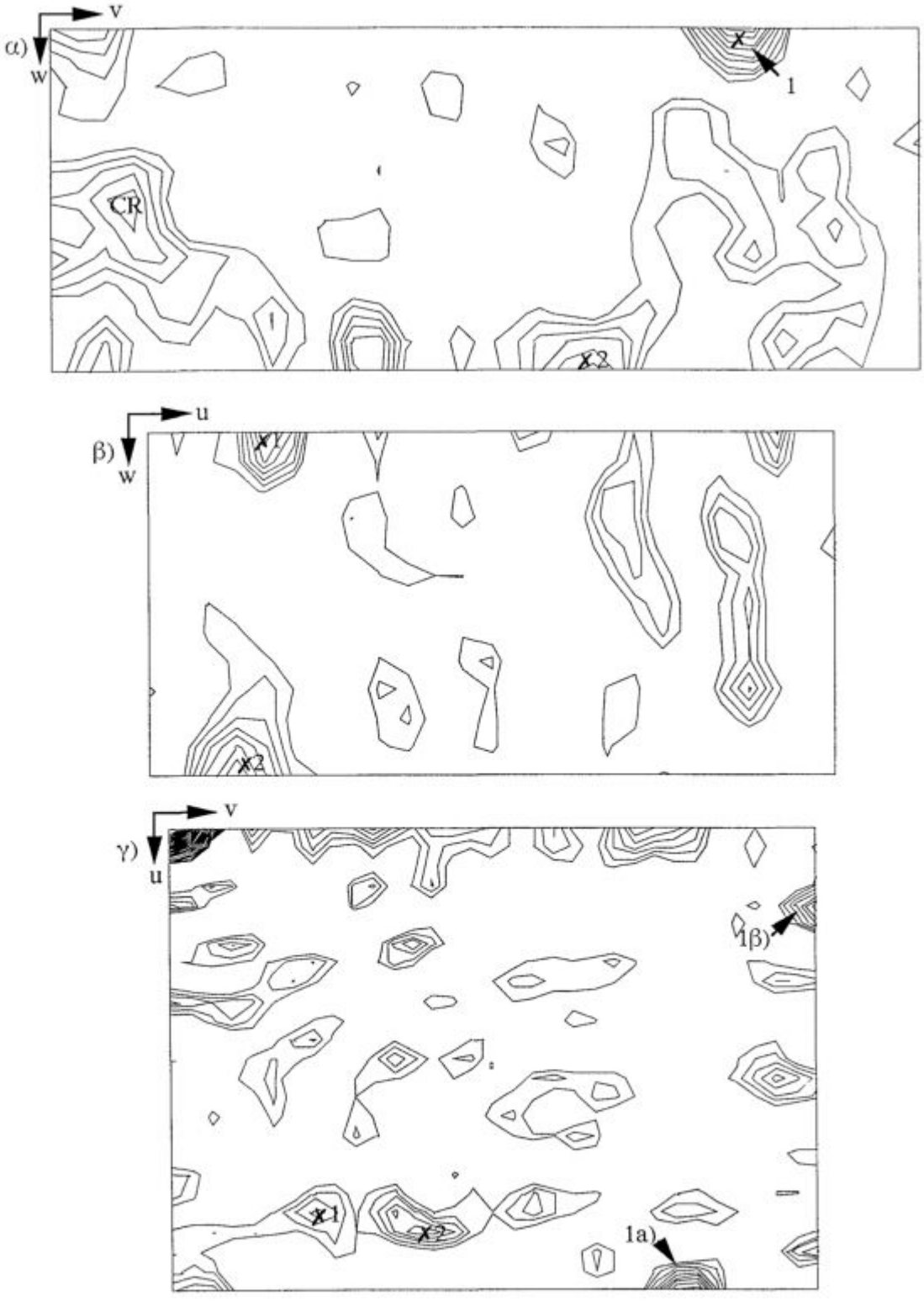
Εικόνα 18. *20-3 Å Συνάρτηση ανώμαλων διαφορών Patterson για το παράγωγο του Πρασεοδυμίου. α) u=1/2 β) v=1/2 γ) w=0. Ισοϋψείς απο το 1σ ανά 1/2σ.1-2 οι κύριες θέσεις πρόσδεσης και 3-4 οι δευτερεύουσες (πίνακας 17).*



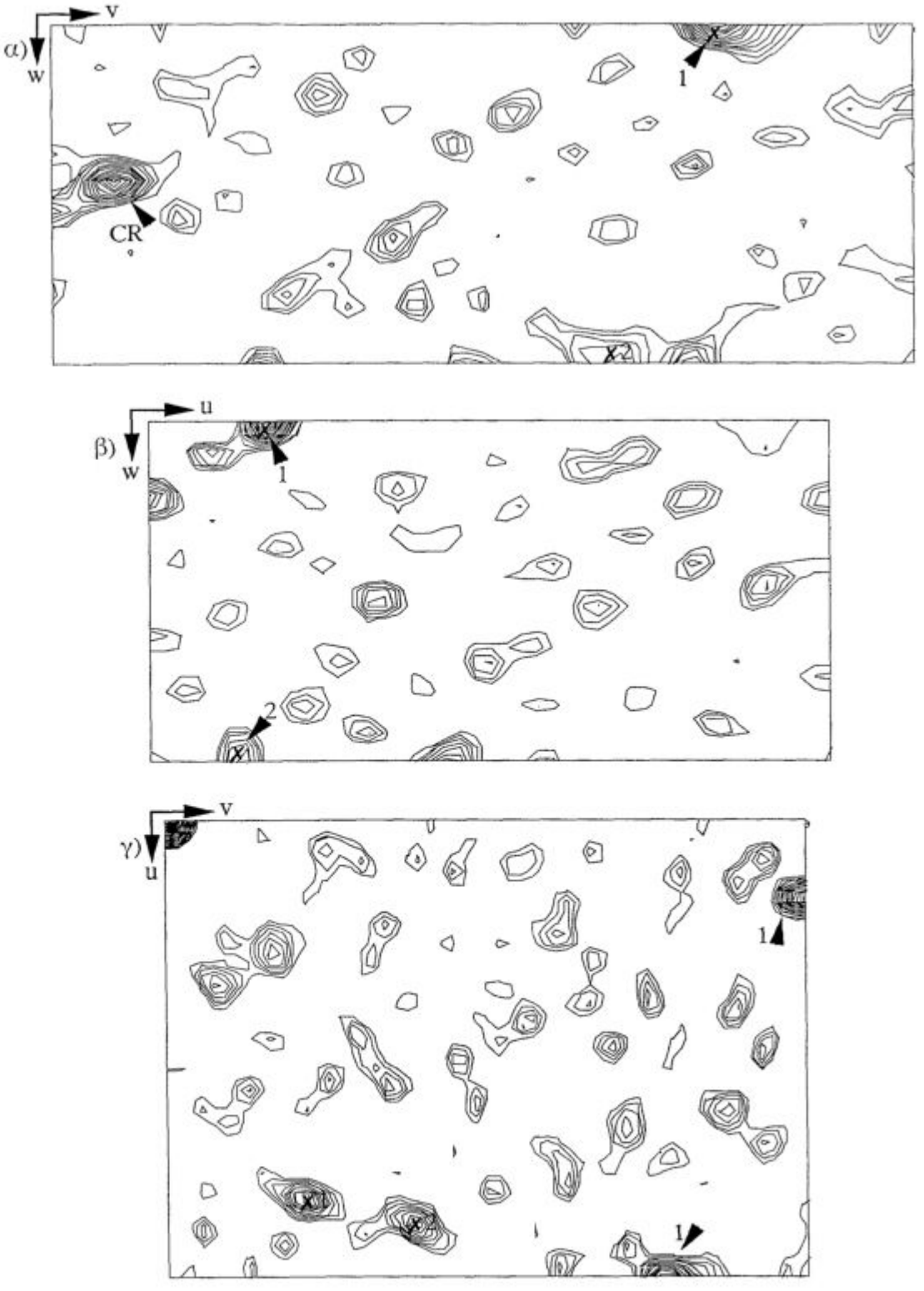
Εικόνα 19. *20-5 Å Συνάρτηση διαφορών Patterson για το παράγωγο του Υττεροβίου. α) u=1/2 β) v=1/2 γ) w=0. 1-2 οι κύριες θέσεις πρόσδεσης και 3-4 οι δευτερεύουσες (πίνακας17)*



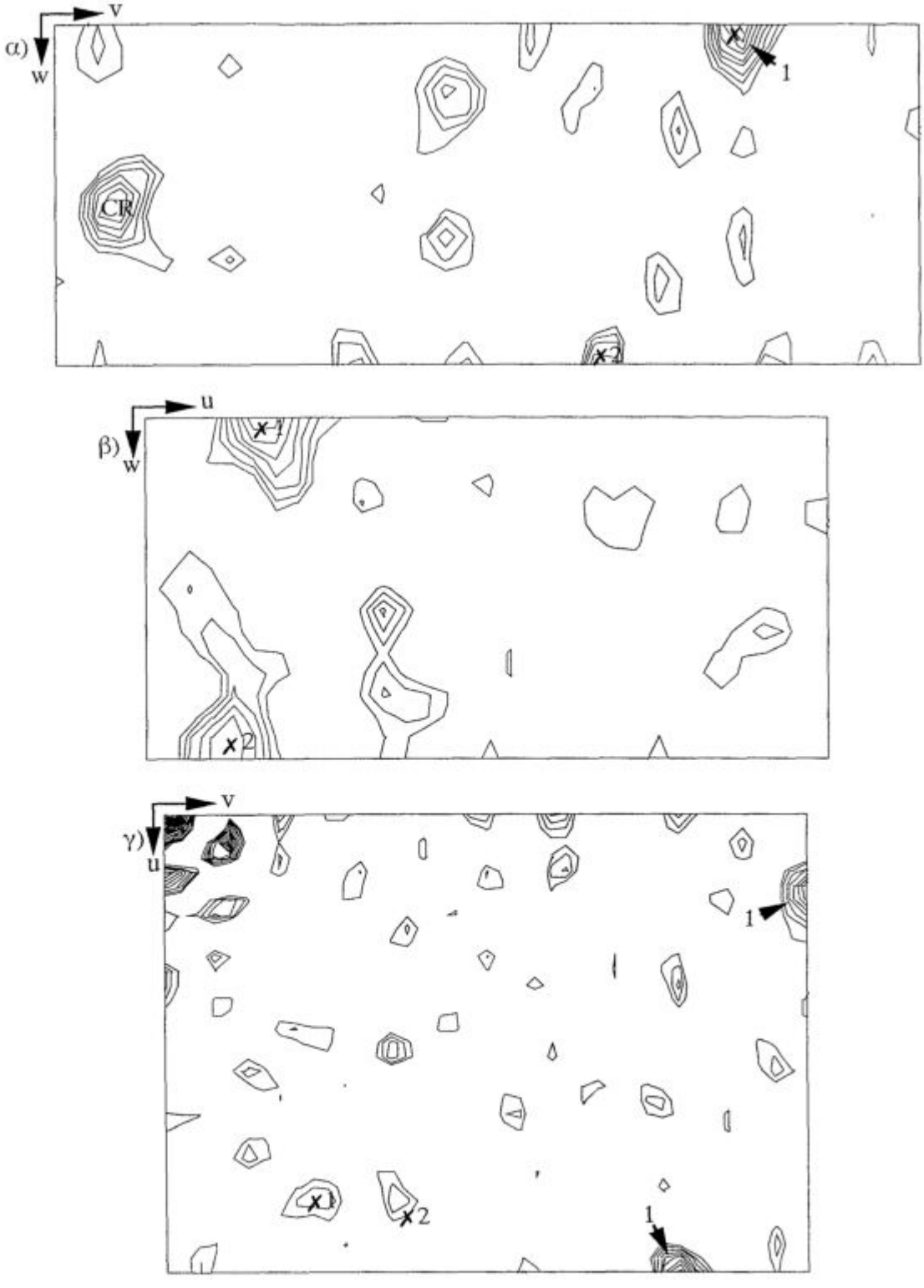
Εικόνα 20. 20-3 Å *Συνάρτηση ανώμαλων διαφορών Patterson για το παράγωγο του Υττερβίου.* *α) u=1/2 β) v=1/2 γ) w=0. Ισοϋψείς απο το 1σ ανά 1/2σ. 1-2 οι κύριες θέσεις πρόσδεσης και 3-4 οι δευτερέυουσες (πίνακας 17).*



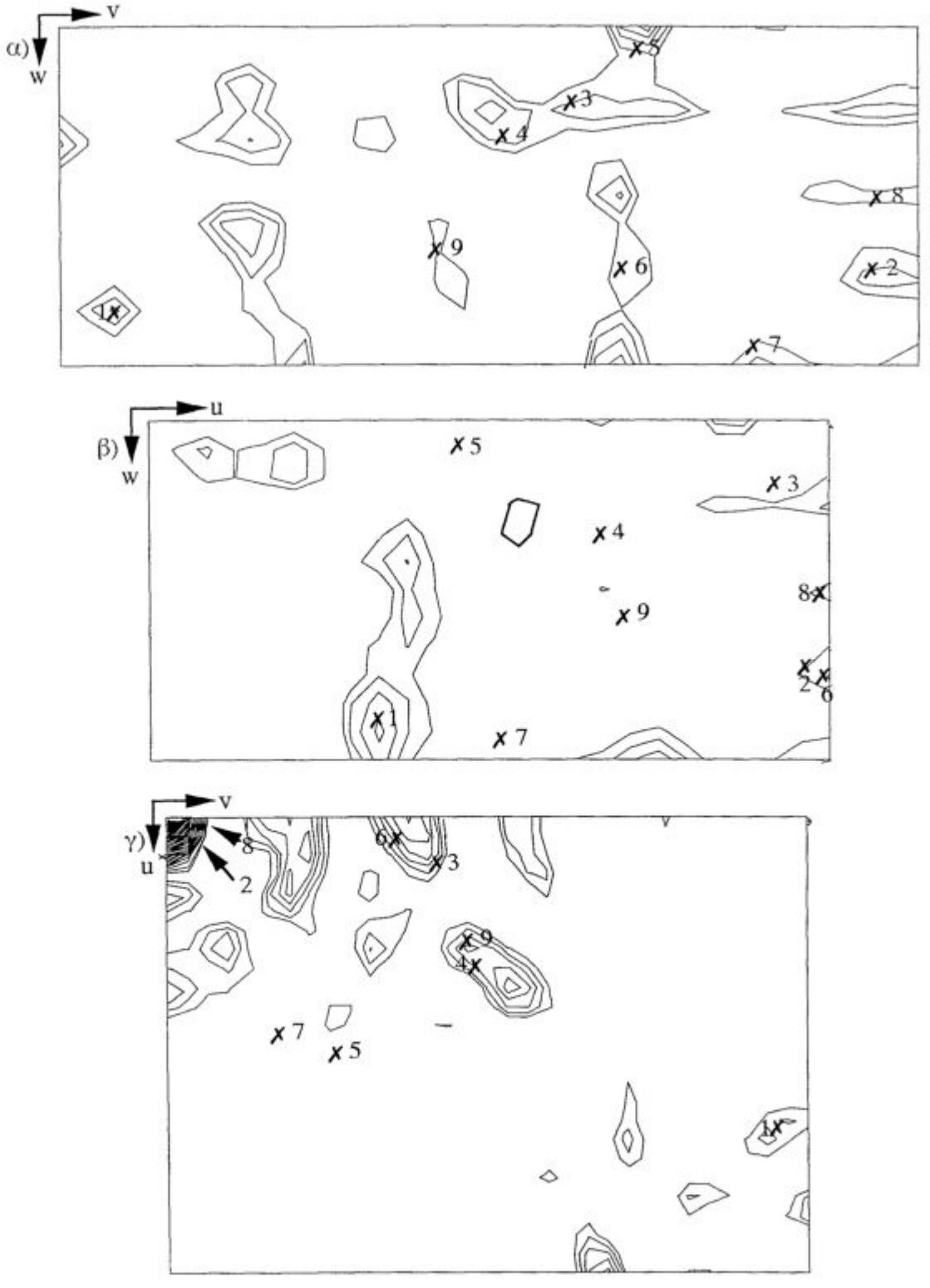
Εικόνα 21. *20-5 Å Συνάρτηση διαφορών Patterson για το παράγωγο του Thiomersal (Hg). α) u=1/2 β) v=1/2 γ) w=0. Ισοϋψείς απο το 1σ ανά 1/2σ. 1-2 οι θέσεις πρόσδεσης (πίνακας 17). CR κορύφή ανύσματος μεταξύ των δύο θέσεων Hg. Και οι τρείς κορυφές για την θέση 1 λόγω ειδικής θέσεως του ατόμου (Z=1/2) εμφανίζονται στην τομή Harker w=0.*



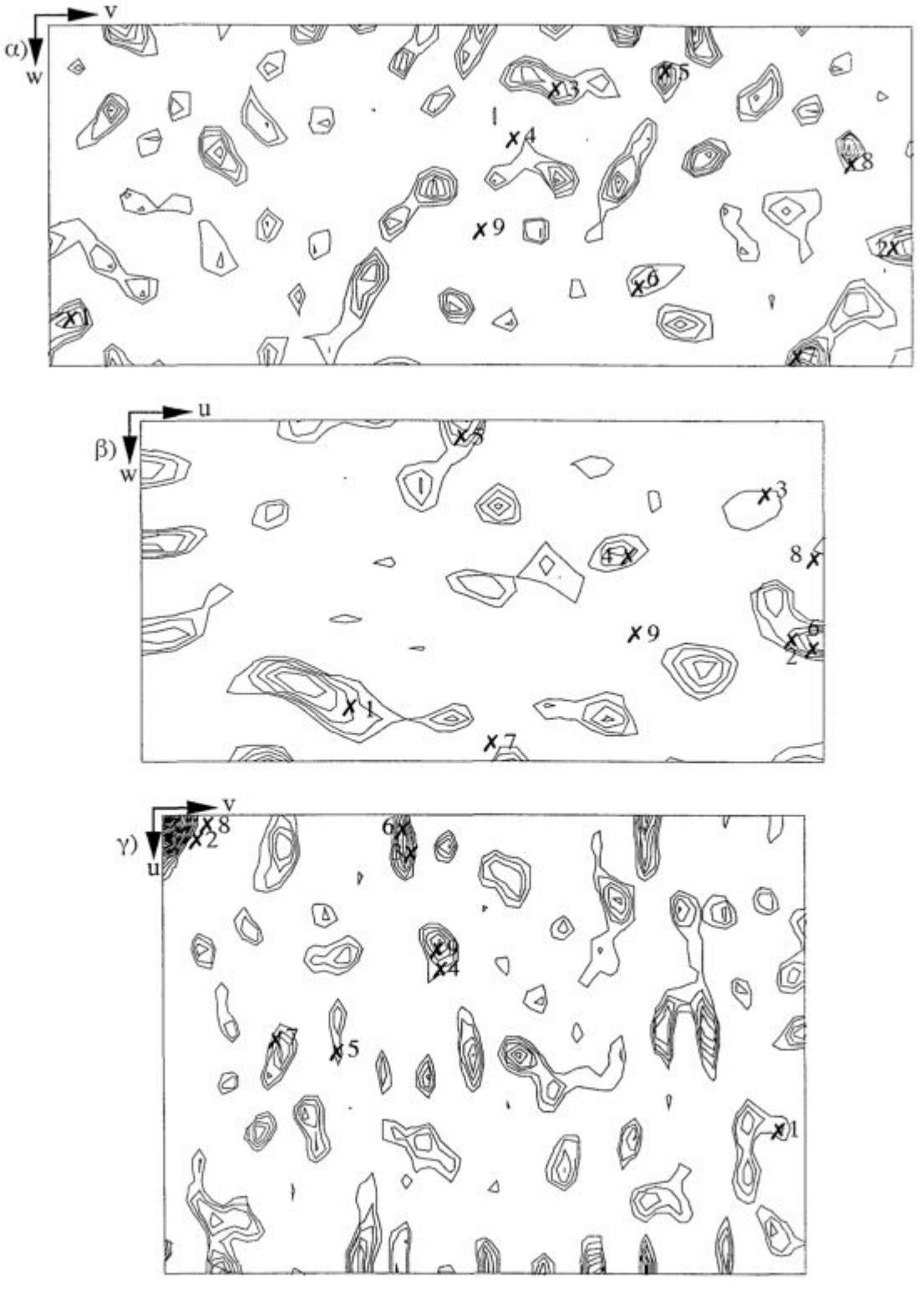
Εικόνα 22. *20-5 Å Συνάρτηση ανώμαλων διαφορών Patterson για το παράγωγο του Thiomersal (Hg). α) u=1/2 β) v=1/2 γ) w=0. Ισουψείς ανά 1/2σ από το 1σ. 1-2 οι θέσεις πρόσδεσης (πίνακας 17).CR κορύφή ανύσματος μεταξύ των δύο θέσεων Hg. Και οι τρείς κορυφές για την θέση 1 λόγω ειδικής θέσεως του ατόμου (Z=1/2) εμφανίζονται στην τομή Harker w=0.*



Εικόνα 23. *20-5 Å Συνάρτηση διαφορών Patterson για το παράγωγο του Thiomersal (Hg1) . α) u=1/2 β) v=1/2 γ) w=0. Ισουψείς ανά 1/2σ από το 1σ. 1-2 οι θέσεις πρόσδεσης (πίνακας 17).CR κορύφή ανύσματος μεταξύ των δύο θέσεων Hg. Και οι τρείς κορυφές για την θέση 1 λόγω ειδικής θέσεως του ατόμου (Z=1/2) εμφανίζονται στην τομή Harker w=0.*



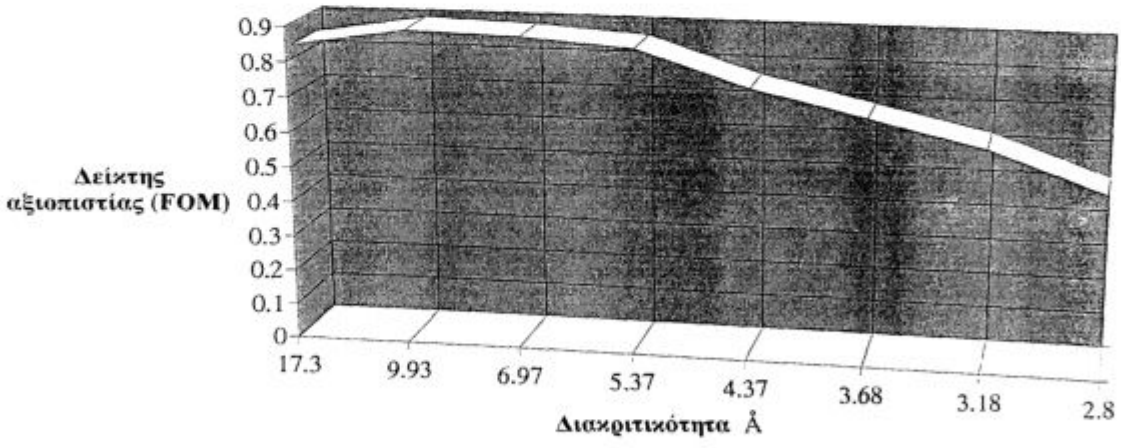
Εικόνα 24. *Επίπεδα Harker απο την 20-5 Å σύνθεση Patterson για τον Αργυρο. α) u=1/2 b) v=1/2 c)w=0. Ισουψείς ανά 1/2σ από το 1σ. 1-9 κύριες και δευτερεύουσες θέσεις πρόσδεσης του Ag.*



***Εικόνα 25**. Επίπεδα Harker απο την 20-3 Å σύνθεση Patterson ανώμαλων διαφορών για τον Αργυρο. α) u=1/2 b) v=1/2 c) w=0. Ισουψείς από το 1σ ανά 1/2σ. 1-9 κύριες και δευτερεύουσες θέσεις πρόσδεσης του Ag.*

Β 7.3. Προσδιορισμός των φάσεων για την φυσική πρωτεΐνη.

Ο προσδιορισμός των φάσεων της φυσικής πρωτεΐνης έγινε στον τελευταίο κύκλο βελτιστοποίησης των παραμέτρων των βαρειών ατόμων με το πρόγραμμα MLPHARE {CCP4,1979}. Η ποιότητα των φάσεων που προέκυψαν περιγράφεται από τον μέσο δείκτη αξιοπιστίας (Figure Of Merit) m=<συν(Δαj)> όπου Δαj το σφάλμα της γωνίας της φάσης για μία ανάκλαση {Blundell & Johnson, 1976}. Ο μέσος δείκτης αξιοπιστίας των φάσεων που προέκυψαν με την μέθοδο πολλαπλής ισόμορφης (MIR) αντικατάστασης για τα δεδομένα μέχρι 2.8Å ήταν <m>=0.59 και μεγαλύτερος από 0.50 για το σύνολο κελυφών διακριτικότητας μέχρι τα 2.8Å (εικόνα 26). Για τον υπολογισμό των χαρτών ηλεκτρονιακής πυκνότητας χρησιμοποιήθηκαν σε σύνθεση Fourier με το πρόγραμμα FFT {CCP4, 1979} οι "βέλτιστες" φάσεις (best phases) και τα πλάτη *Fo* σταθμισμένα με τον δείκτη αξιοπιστίας m. Η ποιότητα των χαρτών ηλεκτρονιακής πυκνότητας που υπολογίστηκαν με τις φάσεις αυτές (εικόνα 27) μας επέτρεψε σε πρώτη παρατήρηση την αναγνώριση χαρακτηριστικών στοιχείων της δευτεροταγούς δομής όπως και πλευρικών αλυσίδων αμινοξικών καταλοίπων. Στο σημείο αυτό η παρουσία δεξιόστροφων α-ελίκων στον χάρτη, μας έδειξε ότι η επιλογή του εναντιομερούς του μοντέλου των βαρειών ατόμων υπήρξε σωστή.



Εικόνα 26. Μεταβολή του δείκτη αξιοπιστίας των MIR φάσεων αυξανομένης της διακριτικότητας.

Β 7.4. Τροποποίηση πυκνότητας του MIR χάρτη.

Παρά το γεγονός ότι η ποιότητα των φάσεων που προέκυψαν με την MIR μέθοδο υπήρξε ικανοποιητική για τον υπολογισμό χαρτών ηλεκτρονιακής πυκνότητας των οποίων η ερμηνεία δεν παρουσίαζε ιδιαίτερα προβλήματα, έγινε προσπάθεια για την περαιτέρω βελτίωση των φάσεων με τις μεθόδους τροποποίησης της ηλεκτρονιακής πυκνότητας {Tulinsky, 1985} : εξομάλυνσης διαλύτη (solvent flatening) και μοριακής εξισορρόπισης (molecular averaging). Οι φάσεις που προέκυψαν με τίς μεθόδους αυτές αν και εμφάνιζαν αρκετά ψηλότερες τιμές του μέσου δείκτη αξιοπιστίας (≈80%) δεν παρουσίαζαν ανάλογη βελτίωση στην οπτική εικόνα του χάρτη. Σαν αποτέλεσμα η χρήση τους στην κατασκευή του μοντέλου περιορίστηκε στην επιβεβαίωση επιλογών στα πιό αμφιλεγόμενα στην ερμηνεία τους σημεία του MIR χάρτη. Η όχι σημαντική βελτίωση των χαρτών ύστερα από την εφαρμογή των τεχνικών τροποποίησης του χάρτη μπορεί να ερμηνευτεί αν λάβουμε υπόψη μας ότι η ποιότητα του MIR χάρτη υπήρξε πολύ καλή στο μεγαλύτερο μέρος του, ενώ στις περιοχές που αυτή δεν ήταν ικανοποιητική η μετέπειτα εργασία έδειξε ότι πρόκειται για περιοχές βρόγχων όπου η πρωτεΐνη παρουσιάζει έλλειψη τάξης. Κατά συνέπεια η ασάφεια του χάρτη δεν οφείλετο στην ποιότητα των φάσεων και δεν υπόκειται σε σημαντική βελτίωση από τις τεχνικές τροποποίησης της πυκνότητας. Τα πρωτόκολλα που χρησιμοποιήθηκαν και για τις δύο μεθόδους περιγράφονται στα συνοδευτικά κείμενα του πακέτου προγραμμάτων CCP4 {CCP4, 1979}.

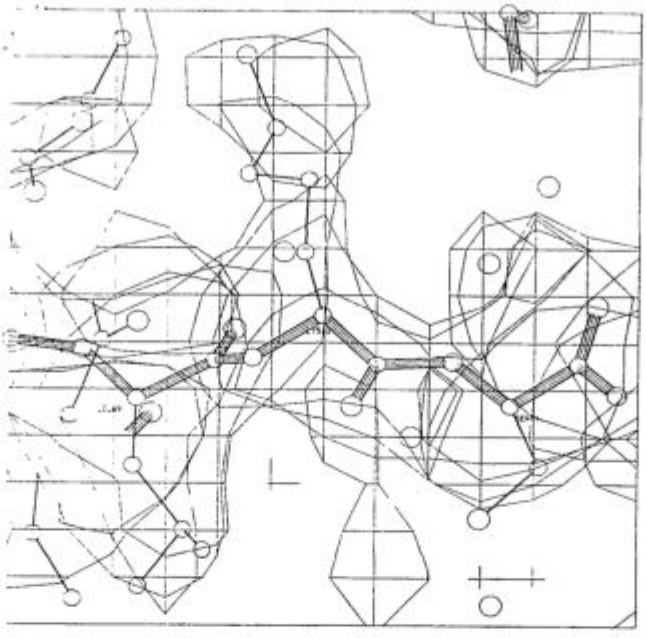
Β 7.4.1 Εξομάλυνση διαλύτη.

Η μέθοδος εξομάλυνσης διαλύτη βασίζεται στην ιδέα ότι στις περιοχές του χάρτη που αντιστοιχούν σε μή διατεταγμένα μόρια διαλύτη η ηλεκτρονιακή πυκνότητα πρέπει να είναι ομοιόμορφη και χαμηλή. Στην περίπτωση των κρυστάλλων της R.*Pvu*II η περιεκτικότητα σε διαλύτη είναι ≈57% και πράγματι μεγάλες περιοχές του MIR χάρτη εμφανίζονται οπτικά ως μη πρωτεϊνικές ("άδειες" από πυκνότητα). Για την εφαρμογή της μεθόδου απαραίτητος είναι ο προσδιορισμός των ορίων μεταξύ της πρωτεΐνης και του διαλύτη. Η όλη διαδικασία πραγματοποιείται αυτοματοποιημένα από σειρά προγραμμάτων του CCP4 πακέτου (Δ 5.7.) και αποτελεί εφαρμογή της μεθόδου όπως αναπτύχθηκε από τον B.C. Wang {Wang, 1985α}. Το ποσοστό του περιεχομένου της ασύμμετρης μονάδας που χρησιμοποιήθηκε ήταν στους πρώτους κύκλους της διαδικασίας 44%. Μικρότερο από το πραγματικό για την αποφυγή της ενδεχόμενης χρήσης από το πρόγραμμα, περιοχών της πρωτεΐνης, ως περιοχών διαλύτη. Σε επόμενους κύκλους και αφού είχε επέλθει βελτίωση του μέσου δείκτη αξιοπιστίας των

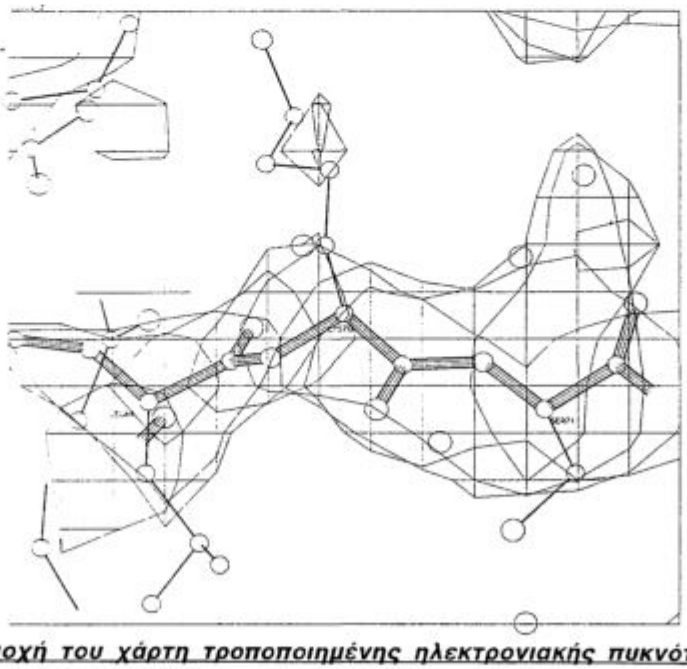
φάσεων ανεβάσαμε το ποσοστό διαλύτη στο 50%. Η φάσεις από την τροποποίηση της ηλεκτρονιακής πυκνότητας κάθε κύκλου συνδυάστηκαν με τις MIR φάσεις με το πρόγραμμα Sigmaa {CCP4, 1979}, {Read, 1986}. Η τελικές φάσεις ύστερα από 8 κύκλους εξομάλυνσης διαλύτη είχαν μέσο δείκτη αξιοπιστίας 80%, ο τροποποιημένος χάρτης παρουσίαζε καλύτερη ποιότητα στις περιοχές της κύριας αλυσίδας σε σχέση με τον MIR, αντίθετα η πυκνότητα σε πλευρικές αλυσίδες αμινοξέων εμφανιζόταν εξασθενημένη (εικόνα 28).

Β 7.4.2. Μοριακή εξισσορρόπιση.

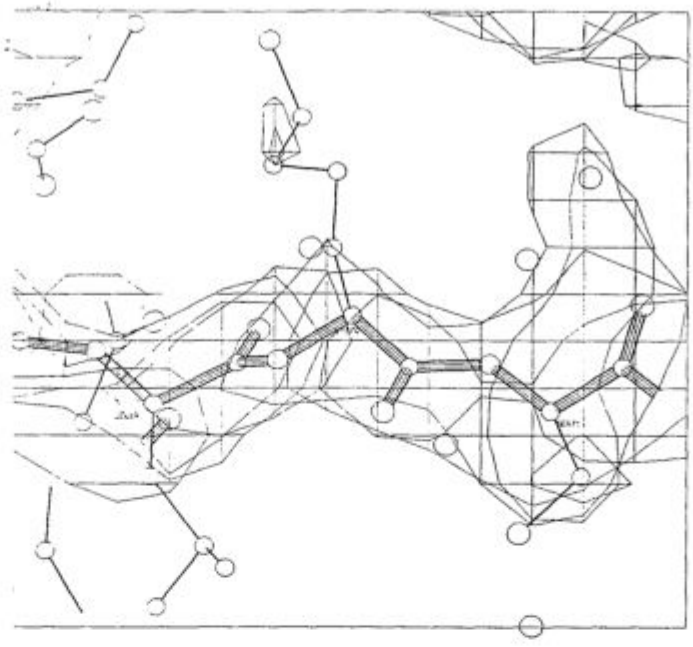
Η τεχνική της "μοριακής εξισορρόπισης" βασίζεται στην παρουσία δύο μονομερών συνδεόμενων με την μή κρυσταλλογραφική συμμετρία στην ασύμμετρη μονάδα {Bricogne in CCP4, 1979}, {Bricogne, 1976}. Αυτό έχει ως αποτέλεσμα η πληροφορία για την δομή του μονομερούς να βρίσκεται δύο φορές στα δεδομένα των ανακλάσεων και φυσικά στους χάρτες ηλεκτρονιακής πυκνότητας. Προυπόθεση για την χρήση της πληροφορίας αυτής είναι η ακριβής γνώση της συμμετρίας που συνδέει τα δύο μονομερή. Στην συγκεκριμένη περίπτωση η πληροφορία για την θέση και την διεύθυνση του άξονα προήλθε αρχικά από τις θέσεις των βαρειών ατόμων στα παράγωγα, και στην συνέχεια, αφού έγινε η αρχική ανεξάρτητη μοντελοποίηση μέρους των δύο μονομερών, προσδιορίστηκε με μεγαλύτερη ακρίβεια κατευθείαν από το μοντέλο της πρωτεΐνης (πρόγραμμα SUPPOS {CCP4, 1979}). Η χρήση της μεθόδου αυτής ξεκινάει με την παραδοχή ότι ο μη κρυσταλλογραφικός άξονας είναι ακριβής και ισχύει για όλες τις περιοχές της πρωτεΐνης με τις ίδιες παραμέτρους, γεγονός το οποίο δεν ισχύει σε αρκετές περιπτώσεις πρωτεϊνών. Ειδικότερα για την συγκεκριμένη περίπτωση το ενδεχόμενο της μή ακρίβειας του άξονα έπρεπε να ληφθεί σοβαρά υπόψιν καθώς στην περίπτωση της R.*Eco*RV παρατηρήθηκαν σημαντικές αποκλίσεις από την ακριβή τάξεως-2 συμμετρία με αποτέλεσμα την αποτυχία ερμηνείας χάρτη βασισμένου σε τροποποίηση πυκνότητας με την τεχνική της μοριακής εξισορρόπισης. Ετσι το αποτέλεσμα της μεθόδου αντιμετωπίστηκε με επιφύλαξη και χρησιμοποιήθηκε μόνο για συγκεκριμένες περιοχές σε συνδυασμό με τον MIR χάρτη. Η διαδιακασία της μοριακής εξισορρόπισης πραγματοποιήθηκε πάνω στον χάρτη που προέκυψε από την εξομάλυνση διαλύτη με το πρόγραμμα Skewplanes {CCP4, 1979} και μέρος του χάρτη διακρίνεται στην εικόνα 29. Για λόγους συγκριτικούς παρατίθεται προθύστερα η ίδια περιοχή του χάρτη με συντελεστές $2F_O$-F_C (εικόνα 30) που προέκυψε ύστερα από την διαδικασία βελτιστοποίησης του μοντέλου της R.*Pvu*II.



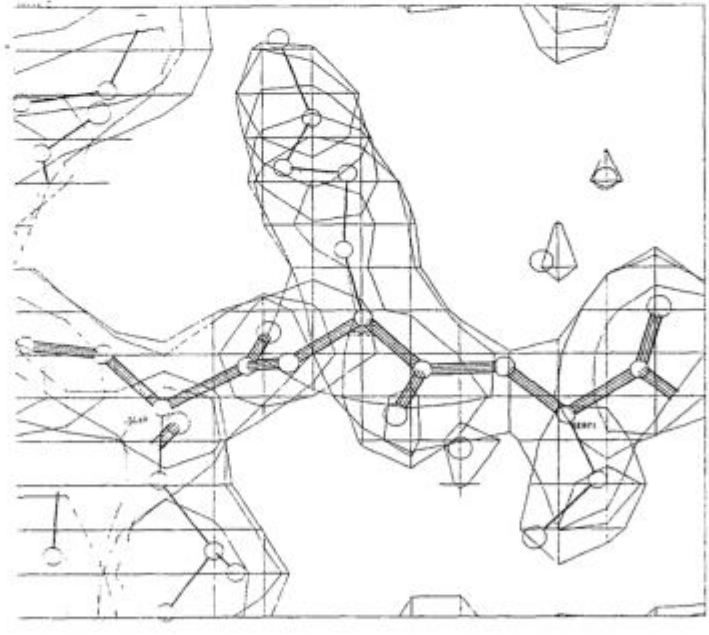
Εικόνα 27. Περιοχή του MIR χάρτη ηλεκτρονιακής πυκνότητας. *Παρουσιάζεται περιοχή (κατάλοιπα Leu69, Lys70 και Ser71) του βc κλώνου που ανήκει στην κεντρική β-πτυχωτή επιφάνεια της R.PvuII. Εντονα σκιασμένη η κύρια αλυσίδα του μοντέλου. Οι ισουψείς δείχνουν περιοχές πυκνότητας μεγαλύτερης από το 1σ.*



Εικόνα 28. Περιοχή του χάρτη τροποποιημένης ηλεκτρονιακής πυκνότητας με την τεχνική Εξομάλυνσης Διαλύτη. *Ο χάρτης παρουσιάζει την ίδια περιοχή με αυτήν της εικόνας 27.*



Εικόνα 29. Περιοχή του χάρτη ηλεκτρονιακής πυκνότητας ύστερα από Εξομάλυνση Διαλύτη και Μοριακή Εξισσορόπιση. *Παρουσιάζεται η ίδια περιοχή του χάρτη με αυτήν της εικόνας 27.*



Εικόνα 30. Περιοχή του χάρτη ηλεκτρονιακής πυκνότητας με συντελεστές $2F_o$-F_c. *Ο χάρτης παρουσιάζει την ίδια περιοχή με αυτήν της εικόνας 27.*

Β 8. Κατασκευή του μοντέλου της R.*Pvu*II.

Το χτίσιμο του μοντέλου πραγματοποιήθηκε πάνω σε MIR χάρτη υπολογισμένο με συντελεστές m*F* και α_{best} φάσεις σε διακριτικότητα 2.8Å. Χάρτες τροποποιημένοι με τις τεχνικές εξομάλυνσης διαλύτη και μοριακής εξισορρόπισης χρησιμοποιήθηκαν ως βοηθήματα στα σημεία όπου ο MIR χάρτης δεν ήταν ικανοποιητικά ερμηνεύσιμος (εικόνες 27-29).

Β 8.1. Ο σκελετός και το αρχικό μοντέλο.

MIR χάρτης υπολογισμένος για ολόκληρη την μοναδιαία κυψελίδα χρησιμοποιήθηκε για την αυτόματη δημιουργία ενός πρώτου σκελετού των C^{α} ατόμων της πρωτεΐνης {Greer, 1985} με το πρόγραμμα ABONES {Jones & Thirup, 1986}, συνοδευτικού προγράμματος του "FRODO" {Jones, 1988} και μετατράπηκαν για χρήση με το πρόγραμμα "O" {Jones & Kjeldgaard 1992}, {Jones, Zou & Cowan, 1991}. Με βάση τον σκελετό αυτό καθορίστηκαν τα όρια του διμερούς μέσα στην στοιχειώδη κυψελίδα και έγινε επέκταση του χάρτη με το πρόγραμμα EXTEND {CCP4, 1979} σε περιοχή που να περιλαμβάνει ένα πλήρες διμερές της R.*Pvu*II.

Η αναγνώριση στον MIR χάρτη στοιχείων της δευτεροταγούς δομής της πρωτεΐνης, κυρίως α-ελίκων, σε συνδυασμό με τις προβλέψεις της δευτεροταγούς δομής από την αλληλουχία, με το πρόγραμμα PREDICT {Eliopoulos et al., 1982}, διευκόλυνε την αναγνώριση πάνω στον χάρτη περιοχών της αλληλουχίας της πρωτεΐνης. Μία προβλεπόμενη αμινοτερματική έλικα (Asp5-Lys25) αναγνωρίστηκε σε πρώτη φάση και χαρακτηριστικά των πλευρικών αλυσίδων όπως αυτά προβλέπονταν από την αλληλουχία ταυτοποιήθηκαν με σχετική ευκολία διευκολύνοντας τις επιλογές σε αμφιλεγόμενα σημεία του χάρτη {Richardson & Richardson, 1985}. Ενδιαφέρον πρόβλημα στο σημείο αυτό υπήρξε το γεγονός ότι η προαναφερθείσα α-έλικα για κάθε μονομερές βρίσκεται τοπολογικά στο εσωτερικό του δεύτερου και είναι συνδεδεμένη με το μονομερές στο οποίο ανήκει με ένα βρόγχο μεγάλων διαστάσεων (εικόνα 39). Η βελτίωση του αρχικού σκελετού έγινε με οπτική επίβλεψη του χάρτη και με τα ειδικά για χειρισμό πρωτεϊνικών σκελετών υποπρογράμματα του "O".

Το βελτιωμένο μοντέλο του σκελετού (C^{α} άτομα μόνο) της πρωτεΐνης μεταφέρθηκε στο πρόγραμμα "FRODO" {Jones, 1988} όπου πραγματοποιήθηκε το χτίσιμο του μοντέλου. Ακολουθώντας τον σκελετό έγινε η μοντελοποίηση της κύριας αλυσίδας των αναγνωρίσιμων στοιχείων δευτεροταγούς δομής, αρχικά χρησιμοποιώντας ιδανικές παραμέτρους φ και ψ γωνιών, και στην συνέχεια προσαρμόστηκε στον χάρτη. Ακολούθησε η μοντελοποίηση των βρόγχων που

συνδέουν τα στοιχεία της δευτεροταγούς δομής, για να προστεθούν τέλος οι πλευρικές ομάδες των καταλοίπων.

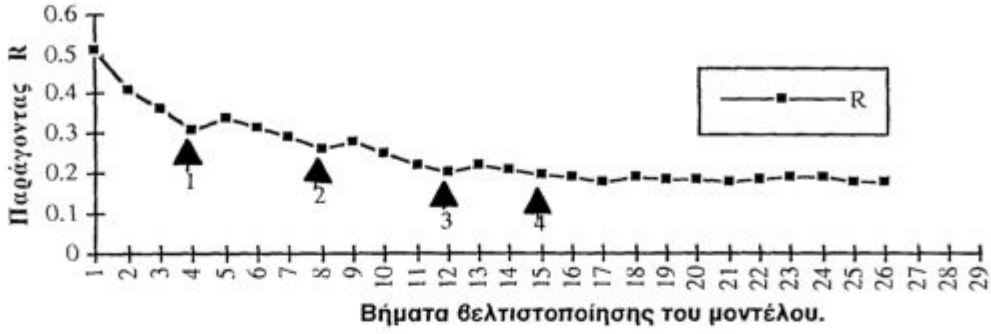
Στον MIR χάρτη ηλεκτρονιακής πυκνότητας μπόρεσαν να τοποθετηθούν συνολικά 141 από τα 156 αμινοξέα του μονομερούς της ώριμης πρωτεΐνης. Τα αμινοξέα Thr49-Glu55 και Trp131-Asn140 τα οποία απουσίαζαν από το μοντέλο που προέκυψε, ανήκουν στους δύο μεγαλύτερους βρόγχους της πρωτεΐνης στην περιοχή των οποίων η ηλεκτρονιακή πυκνότητα του MIR χάρτη υπήρξε ασαφής και δεν επέτρεψε την χωρίς αμφιβολία μοντελοποίηση τους. Δημιουργία του μοντέλου του διμερούς έγινε με εφαρμογή της μη κρυσταλλογραφικής συμμετρίας και ακολούθησε ανεξάρτητη προσαρμογή του μοντέλου του δεύτερου στην ηλεκτρονιακή πυκνότητα. Το μοντέλο του διμερούς που προέκυψε χρησιμοποιήθηκε στην διαδικασία της βελτιστοποίησης που περιγράφεται στην συνέχεια.

Β 8.2. Βελτιστοποίηση του μοντέλου της R.*Pvu*II.

Ακολούθησε βελτιστοποίηση του μοντέλου με την μέθοδο "simulated annealing" {Brunger, Kuriyan & Karplus, 1987} με την βοήθεια του προγράμματος XPLOR {Brunger, 1990} ακολουθώντας το πρωτόκολλο "βραδείας ψύξης" (Slow-Cooling) {Brunger, Krukowski & Erickson, 1990}. Στην προσομοίωση χρησιμοποιήθηκαν ως επιπλέον περιορισμοί αρχικά οι MIR φάσεις (Οπου η φάση της ανάκλασης h περιορίζεται στα όρια $\varphi_{MIR}(h)$ +/- acos(m(h))) και η μή κρυσταλλογραφική συμμετρία. Για την εφαρμογή της μή κρυσταλλογραφικής συμμετρίας έγινε ακριβέστερος προσδιορισμός του πίνακα μετασχηματισμού χρησιμοποιώντας τα ανεξάρτητα μοντελοποιημένα μονομερή. Στην εικόνα 31 συνοψίζεται η πορεία της βελτιστοποίησης.

Σε χάρτες υπολογισμένους με συντελεστές είτε $2F_o$-F_c, είτε με mFo-nFc, όπου οι σταθμίσεις m,n υπολογίστηκαν από το πρόγραμμα SIGMAA {Read, 1986}, τοποθετήθηκαν 13 από τα 15 αμινοξέα που απουσίαζαν από το αρχικό μοντέλο. Τα κατάλοιπα Arg54 και Glu55 απουσιάζουν από το τελικό μοντέλο και αποτελούν τμήμα ευκίνητου βρόγχου, όπως φαίνεται και από τους ψηλούς παράγοντες θερμοκρασίας B των αμινοξικών καταλοίπων που προηγούνται και έπονται στην αλληλουχία Gly53 και Gly56 (εικόνα 36). Στα τελικά στάδια της βελτιστοποίησης έγινε προσθήκη μορίων νερού στο μοντέλο. Συνολικά προστέθηκαν 79 μόρια νερού με ένα η περισσότερους πιθανούς υδρογονικούς δεσμούς με πρωτεϊνικά άτομα. Συνολικά δημιουργούνται 52 υδρογονικοί δεσμοί με άτομα της κύριας αλυσίδας (H_2O <B>=54.9Å^3) και 86 με άτομα των πλευρικών αλυσίδων (H_2O <B>=64Å^3). Το μοντέλο που προέκυψε κατατέθηκε στην βάση δεδομένων Brookhaven και ο κωδικός της καταχώρησης είναι «1Pvu».

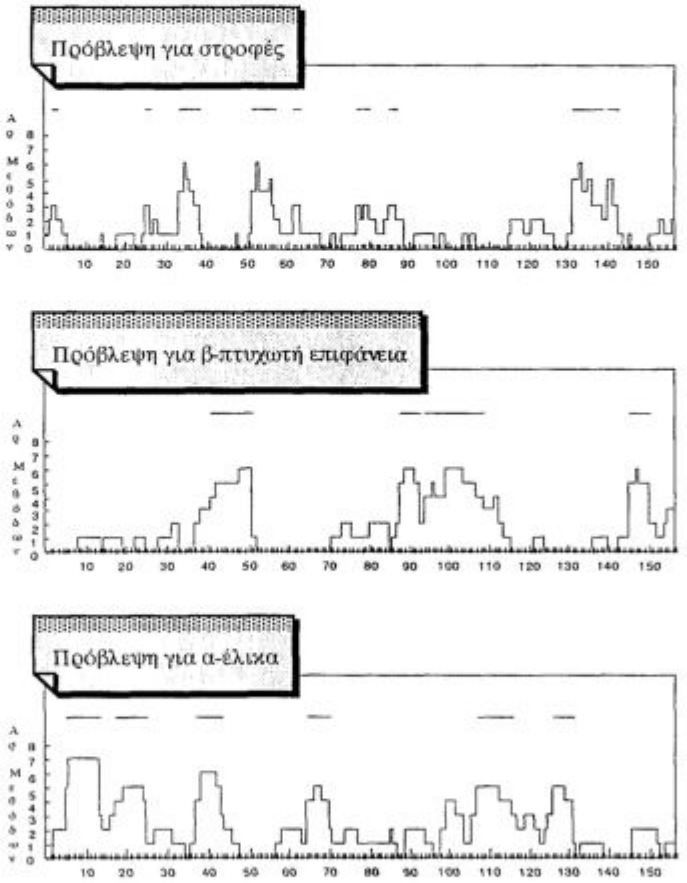
Πορεία της βελτιστοποίησης του μοντέλου της R.PvuII



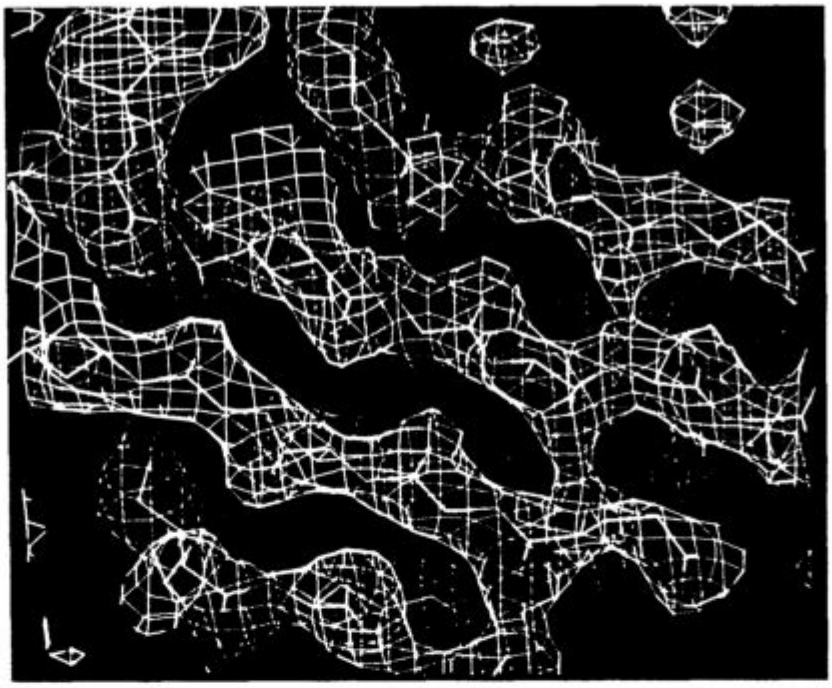
Εικόνα 31. Βελτιστοποίηση του μοντέλου που προέκυψε από τον MIR χάρτη. Στο διάγραμμα σημείωνονται με βέλη οι τιμές R στο τέλος των κυριώτερων σταδίων της βελτιστοποίησης. Πραγματοποιήθηκαν 3 στάδια βελτιστοποίησης βασισμένα στη μέθοδο simulated annealing καθένα από τα οποία περιελάμβανε τα βήματα α) 50-100 κύκλους ελαχιστοποίησης ενέργειας του μοντέλου (συμπεριλαμβανομένου του κρυσταλλογραφικού όρου), β) Εκκίνηση από αρχική θερμοκρασία 4000°K (στάδια 1,2) ή 1000°K (στάδιο 3) και βραδεία ψύξη στους 300°K γ) Νέα ελαχιστοποίηση ενέργειας (50-85 κύκλοι) και δ) Βελτιστοποίηση του παράγοντα θερμοκρασίας B. Στα πρώτα 2 στάδια χρησιμοποιήθηκαν ως επιπλέον περιορισμοί (restraints) οι MIR φάσεις και η μή κρυσταλλογραφική συμμετρία ενώ στο στάδιο 3 χρησιμοποιήθηκε μόνο η μή κρυσταλλογραφική συμμετρία. Τέλος πραγματοποιήθηκε σειρά στάδιων που περιελάμβανε βελτιστοποίηση (20-60 κύκλοι) με την μέθοδο συζηγούς διαβάθμισης (Conjugate gradient) και βελτιστοποίηση του παράγοντα B με ενδίαμεσες διορθώσεις του μοντέλου ύστερα από οπτική επίβλεψη του χάρτη ηλεκτρονιακής πυκνότητας με τη βοήθεια του προγράμματος FRODO {Jones, 1988}. Στο στάδιο 1 της βελτιστοποίησης χρησιμοποιήθηκαν ανακλάσεις στην περιοχή διακριτικότητας 30-2.5Å, στα στάδια 2,3,4 στην περιοχή 6.0-2.5Å και στα τελικά στάδια 6.0-2.4Å. Ανακλάσεις με $F>3\sigma(F)$ μετείχαν στους υπολογισμούς σε όλα τα στάδια εκτός του σταδίου 1 όπου δεν υπήρξε περιορισμός. Η προσθήκη των καταλοίπων που απουσίαζαν αρχικά από το μοντέλο έγινε μετά το πρώτο στάδιο της βελτιστοποίησης και η προσθήκη στο μοντέλο μορίων νερού έγινε σταδιακά μετά το στάδιο 4. Επιπλέον στα στάδια 1-3 που ως επιπλέον περιορισμός τέθηκε η μη κρυσταλλογραφική συμμετρία αφέθηκαν ελεύθερες από τον περιορισμό οι περιοχές των βρόγχων της πρωτείνης καθώς και η N-τερματική περιοχή (κατάλοιπα 2-4, 27-30, 50-57, 73-79, 83-88, 104-108, 131-146).

Πίνακας 19. Σύνοψη του αποτελέσματος της βελτιστοποίησης και στατιστικά της γεωμετρίας του μοντέλου της R.*Pvu*II που προέκυψε από την βελτιστοποίηση. Τα στατιστικά υπολογίστηκαν με το πρόγραμμα TNT {Tronrud, Ten Eyck & Mattiews, 1987}.

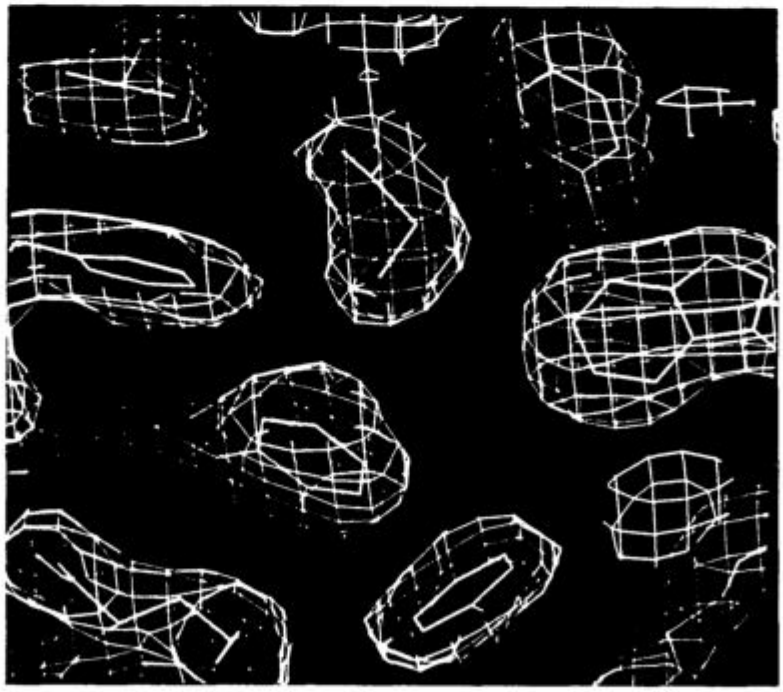
Ορια Διακριτικότητας (Å)	6-2.4
Ανακλάσεις με $F>2\sigma F$	15332
Παράγοντας R (%)	18.4
Πρωτεινικά άτομα του μοντέλου	2544
Μόρια νερού στο μοντέλο	79
R.m.s απόκλιση στις γωνίες δεσμών	3.6 ^{o}
R.m.s. απόκλιση στο μήκος δεσμών	0.023Å
Δίεδρες γωνίες	24.7^{o}
Τριγωνική μη-επιπεδότητα	0.006Å
Επιπεδότητα	0.028Å



***Εικόνα 32. Πρόβλεψη της δευτεροταγούς δομής της R.Pvu*II.** *Πρόγραμμα Predict, μέθοδος Joint. Η πρόβλεψη χρησιμοποιήθηκε για την συσχέτιση της αλληλουχίας με περιοχές του MIR χάρτη ηλεκτρονιακής πυκνότητας.*



Εικόνα 33. Αποψη της κεντρικής β-πτυχωτής επιφάνειας του τελικού 2Fo-Fc χάρτη ηλεκτρονιακής πυκνότητας της R.Pvull .



Εικόνα 34. Αποψη του πυρήνα αρωματικών καταλοίπων του τελικού 2Fo-Fc χάρτη ηλεκτρονιακής πυκνότητας της R.Pvull .

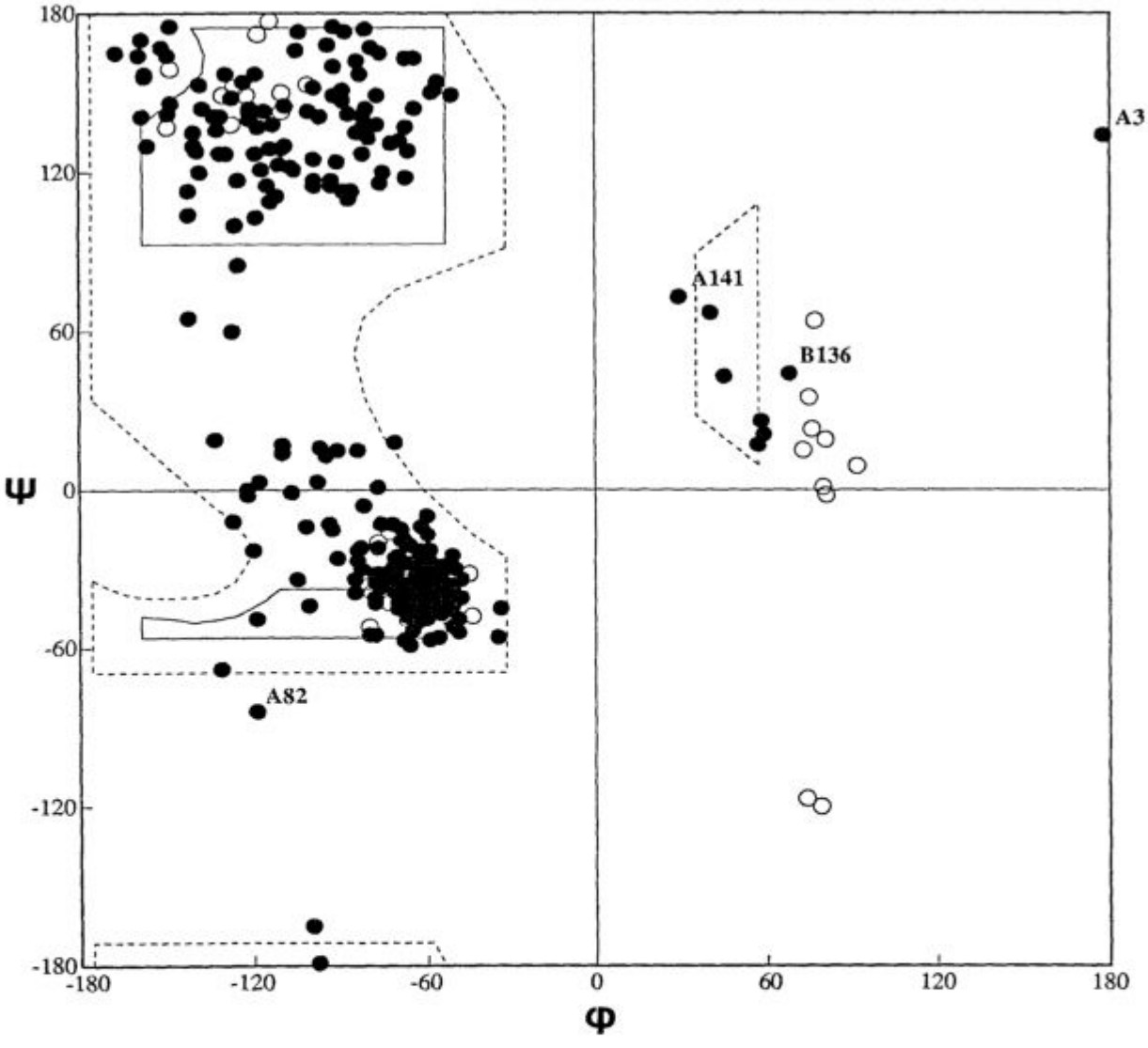
Β 8.3. Ανάλυση της ποιότητας του μοντέλου.

Προκειμένου να εκτιμηθεί η ορθότητα και η ακρίβεια μίας κρυσταλλογραφικά προσδιορισμένης δομής πρωτεΐνης, είναι αναγκαίο να μελετήθει η ποιότητα του μοντέλου πάνω στο οποίο βασίζεται. Ενα ιδανικό μοντέλο πρωτεΐνης χαρακτηρίζεται αφενός από την συμφωνία των γεωμετρικών παραμέτρων του (φ-ψ γωνίες, αποστάσεις δεσμών, γωνίες δεσμών) με τις προβλεπόμενες για πρωτεϊνικά μόρια και αφετέρου από την συμφωνία ανάμεσα στα πειραματικά κρυσταλλογραφικά δεδομένα και αυτά που θα προέκυπταν αν μέσα στον κρύσταλλο βρισκόταν το μοντέλο αντί της πραγματικής πρωτεΐνης. Η σύγκριση αυτή εκφράζεται από τον παράγοντα $R=\Sigma|F_O-F_C|/\Sigma|F_O|$ και τις παραλλαγές του. Ο παράγοντας θερμοκρασίας Β που εκφράζει την κινητικότητα κάθε ατόμου ενός μοντέλου, εν μέρει, μπορεί να αποτελέσει κριτήριο γιά την ποιότητα του μοντέλου σε διαφορετικές περιοχές του μορίου, μπορεί όμως και να εκφράζει μία πραγματική ιδιότητα του μορίου με χρήσιμες προεκτάσεις για την κατανόηση της λειτουργίας του. Στην συνέχεια ακολουθεί μια ανάλυση των ποιοτικών χαρακτηριστικών του μοντέλου της R.*Pvu*II όπως προέκυψε από το τελικό στάδιο της βελτιστοποίησης, προκειμένου να εντοπιστούν ατέλειες του μοντέλου.

<u>Οι γεωμετρικές παράμετροι του μοντέλου.</u>

Στην εικόνα 35 παρουσιάζεται το διάγραμμα Ramachandran για τις φ, ψ γωνίες {Ramachandran & Sasisekharan, 1968} του τελικού μοντέλου της R.*Pvu*II. Το σύνολο σχεδόν των τιμών περιλαμβάνεται στις ιδανικές ή στις επιτρεπτές περιοχές του διαγράμματος. Το κατάλοιπο His136 εντοπίζεται σε περιοχή όπου η ηλεκτρονιακή πυκνότητα είναι ασθενής με πιθανό αποτέλεσμα μία μικρή απόκλιση από την επιτρεπτή περιοχή. Σε αυτό συνηγορεί το γεγονός ότι η His136 του άλλου μονομερούς βρίσκεται στο διάγραμμα μέσα στην γειτονική L-περιοχή. Οι γλυκίνες του μορίου καθώς δεν υπόκεινται σε στερεοχημικούς περιορισμούς βρίσκονται σε περιοχές του διαγράμματος μη επιτρεπτών τιμών για τα υπόλοιπα αμινοξέα. Μικρές αποκλίσεις παρουσιάζονται και για τα κατάλοιπα Thr82 και Asn141 χωρίς όμως να παρατηρούνται και στα δύο μονομερή. Πέρα από τις φ, ψ γωνίες του μοντέλου τα υπόλοιπα στατιστικά της γεωμετρίας του μοντέλου έχουν δοθεί στον πίνακα 19 του κεφαλαίου Β.8.2. χωρίς να υπάρχει κάτι άξιο για σχολιασμό.

Στο διάγραμμα της εικόνας 38 παρουσιάζεται η κατανομή του παράγοντα R με βάση το τελικό μοντέλο ως συνάρτηση της διακριτικότητας. Το rms σφάλμα των συντεταγμένων παραμένει χαμηλότερο των 0.35Å για το σύνολο της διακριτικότητας μέχρι τα 2.4Å.



Εικόνα 35. Σχέδιο Ramachandran για το τελικό μοντέλο της R.Pvull. ***{Biomol System Groningen,1991}*** *Τα κατάλοιπα γλυκίνης εμφανίζονται ως ανοικτοί κύκλοι και όλα τα άλλα κατάλοιπα ως μαύροι κύκλοι. Ονοματισμένα εμφανίζονται τα αμινοξικά κατάλοιπα με φ και ψ γωνίες εκτός των επιτρεπτών περιοχών του διαγράμματος.*

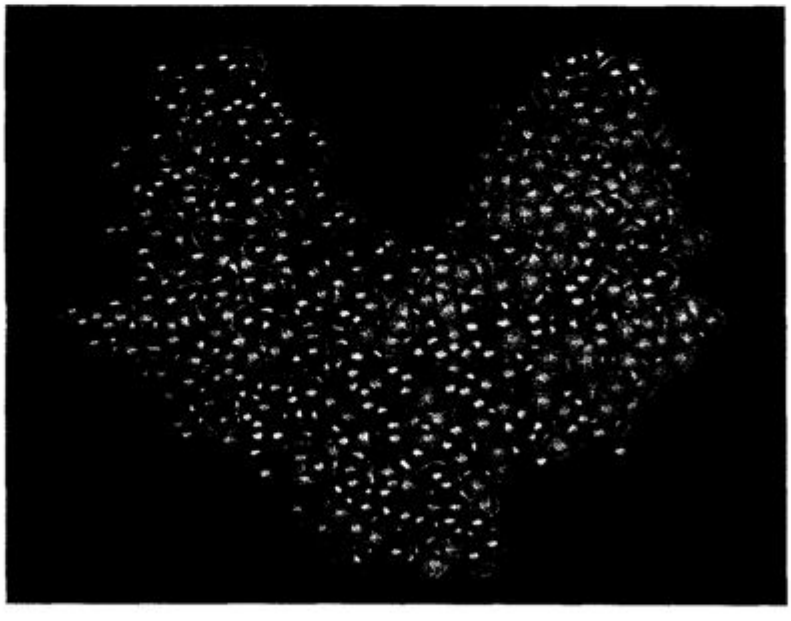
Οι παράγοντες θερμοκρασίας B.

Μελετώντας την κατανομή των παραγόντων θερμοκρασίας B των ατόμων του μορίου της R.*Pvu*II μια πρώτη παρατήρηση είναι η διαφοροποίηση ανάμεσα

στα δύο μονομερή (εικόνα 37). Η διαφοροποίηση αυτή εντοπίζεται κυρίως στις εκτεθιμένες στο διάλυμα περιοχές της πρωτεΐνης (εικόνα 38). Ο μέσος Β παράγοντας (υπολογισμένος με το πρόγραμμα Baverage {CCP4, 1979}) για το μονομερές Α είναι <B>=40.83Å^2 και 46.83Å^2 για τα άτομα της κύριας αλυσίδας και των πλευρικών αλυσίδων αντίστοιχα ενώ για το μονομερές Β <B>=35.69Å^2 (κύρια αλυσίδα) και 41.59Å^2 (πλευρικές αλυσίδες). Η διαφορά αυτή αντανακλά πιθανώς το γεγονός ότι το διαφορετικό κρυσταλλικό περιβάλλον των δύο μονομερών έχει ως συνέπεια διαφορετικές αλληλεπιδράσεις του κάθε μονομερούς με άτομα των γειτονικών διμερών (βλ. κεφάλαιο Β9.3.).

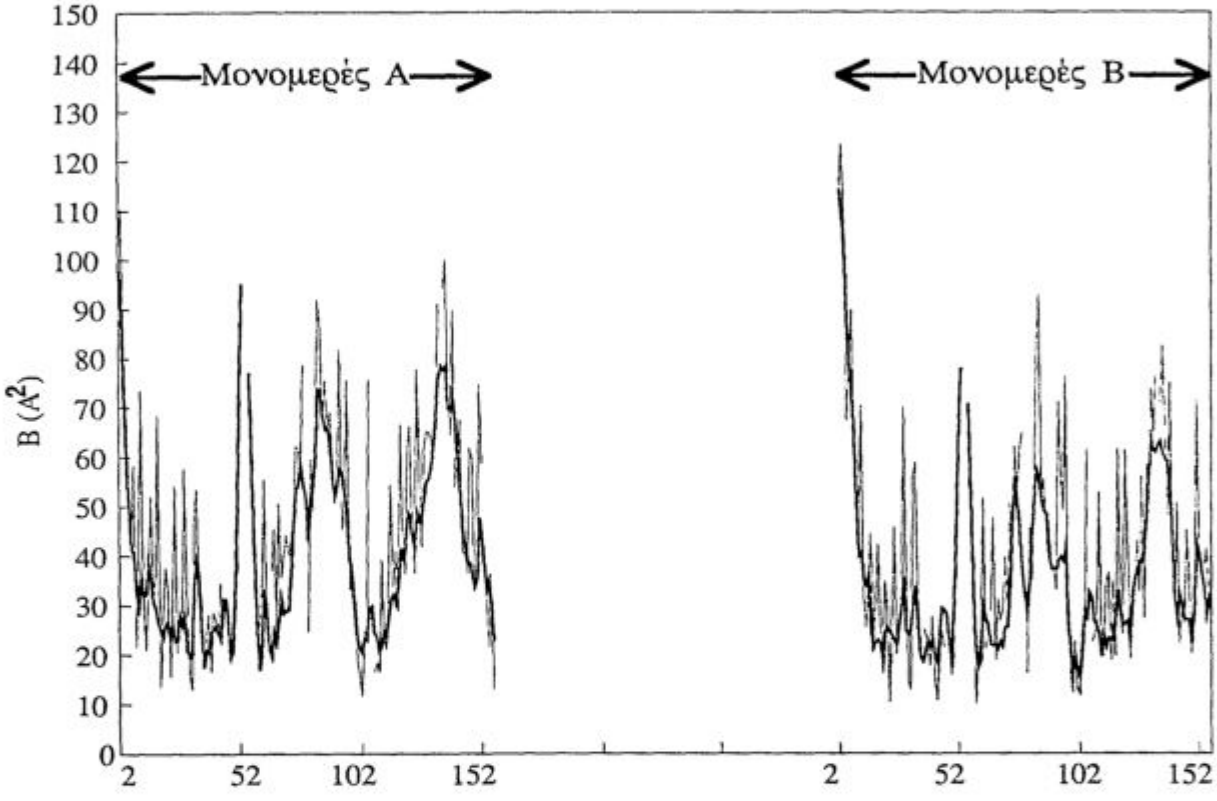
Μέσα σε κάθε μονομερές υπάρχουν τέσσερις περιοχές της πρωτεΐνης (εικόνα 37) που δείχνουν ιδιαίτερη κινητικότητα :

α) Η αμινοτελική περιοχή του μορίου είναι εκτεθειμένη στο διάλυμα και εμφανίζει τον ψηλότερο παράγοντα Β, φαινόμενο συνηθισμένο στις πρωτεϊνικές δομές. Η ευελιξία την αμινοτελικής περιοχής επιβεβαιώνεται απο το γεγονός ότι η ηλεκτρονιακή πυκνότητα στις Ν-τελικές περιοχές των δύο μονομερών διαφοροποιείται με συνέπεια η διαμόρφωση του μοντέλου για τα 3 πρώτα αμινοξέα της πρωτεΐνης (Ser2, His3 και Pro4) να είναι σημαντικά διαφορετική για κάθε μονομερές.



Εικόνα 36. Κατανομή των παραγόντων θερμοκρασίας Β σε χωροπληρωτικό μοντέλο του διμερούς της R.Pvull. *Σε χρωματική διαβάθμιση μπλέ (ελάχιστο)-κόκκινου (μέγιστο) οι παράγοντες Β των ατόμων . Γίνεται φανερή η ανομοιομορφία ανάμεσα στα δύο μονομερή.*

β) Η δεύτερη περιοχή με ψηλούς παράγοντες Β είναι η περιοχή της πρωτεΐνης που περιλαμβάνει τα ευκίνητα κατάλοιπα Gly53 και Gly56 και περικλείει τα κατάλοιπα Arg54 και Glu55 των οποίων η πυκνότητα δεν στάθηκε δυνατόν να γίνει ορατή στους χάρτες ηλεκτρονιακής πυκνότητας. Στην περιοχή αυτή ανήκουν κατάλοιπα που μετέχουν στο καταλυτικό κέντρο της πρωτεΐνης και η περιοχή υφίσταται σημαντικές αλλαγές διαμόρφωσης για την πραγματοποίηση της καταλυτικής λειτουργίας, όπως θα δούμε σε επόμενο κεφάλαιο.

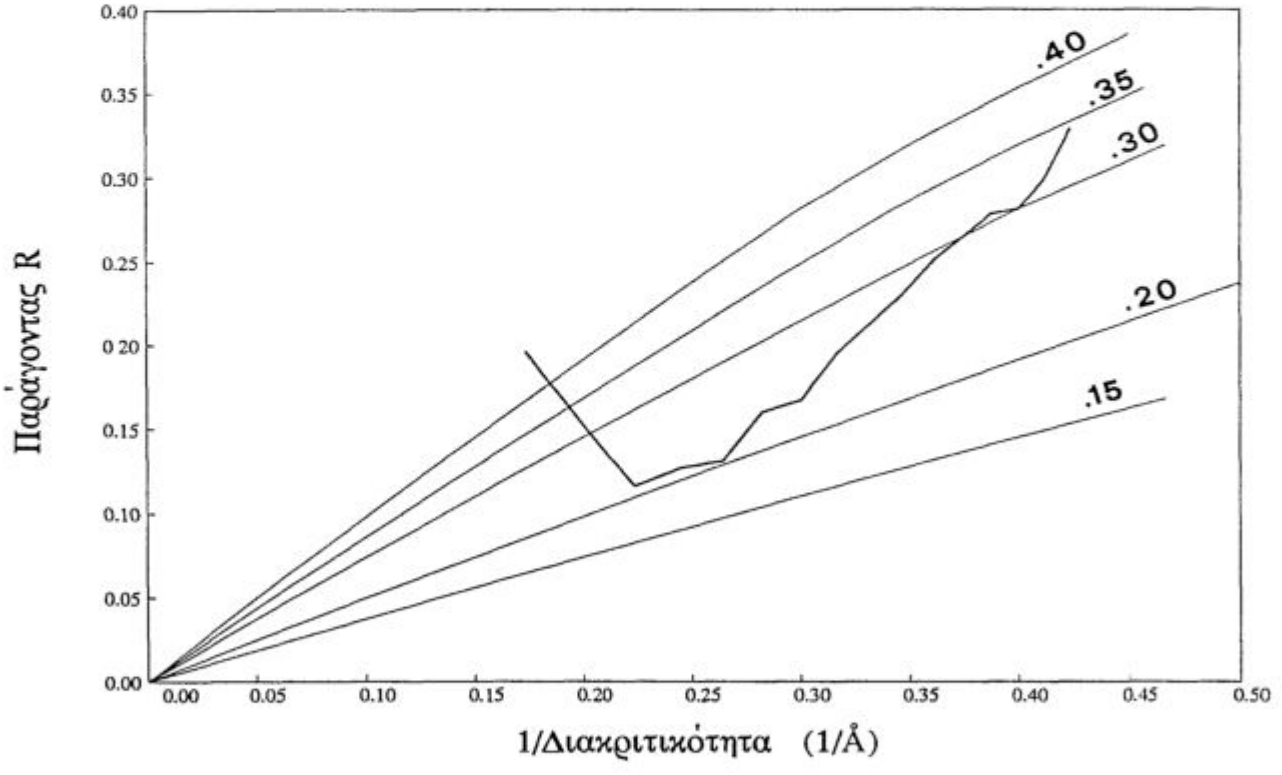


Εικόνα 37. Διάγραμμα της κατανομής των παραγόντων θερμοκρασίας Β κατά μήκος της αλληλουχίας. *Σε ξεχωριστές καμπύλες παρουσιάζονται τα άτομα της κύριας αλυσίδας (παχιά γραμμή) και των πλευρικών αλυσίδων (λεπτή γραμμή). (Biomol System Groningen, 1991)*

γ) Μία δεύτερη περιοχή της πρωτεΐνης που μετέχει στην δημιουργία του καταλυτικού κέντρου, περιλαμβάνει τα κατάλοιπα Asp75, Leu76 και Thr77 και εμφανίζει επίσης ψηλούς Β-παράγοντες.

δ) Η περιοχή των τριών διαδοχικών His83-His85, όπως θα δούμε στο κεφάλαιο B10.2. μετέχει στον σχηματισμό των ειδικών αλληλεπιδράσεων με τις βάσεις του DNA και στην ελεύθερη πρωτεΐνη εμφανίζει σχετικά ψηλούς B-παράγοντες.
ε) Τέλος B-παράγοντες ψηλότερους από 50Å^2 παρουσιάζει και η περιοχή, του βρόγχου μεταξύ της α-έλικας αD και του βg κλώνου, Ser133 - Lys137.

Συνοψίζοντας, όλες οι περιοχές πού εμφανίζουν ψηλούς παράγοντες B ανήκουν σε βρόγχους μεταξύ δευτεροταγών στοιχείων της πρωτεΐνης και εκπληρώνουν κάποιο λειτουργικό ρόλο, είτε αυτός σχετίζεται με την αναγνώριση και την δέσμευση του DNA, είτε με την καταλυτική λειτουργία του μορίου. Παρόμοια κατανομή, των B- παραγόντων, παρατηρήθηκε για την R.*Eco*RV όπου βρόγχοι οι οποίοι μετέχουν στην αναγνώριση και την κατάλυση, εμφανίζουν ιδιαίτερη κινητικότητα στην ελεύθερη πρωτεΐνη {Winkler, 1993}. Εξαίρεση στην R.*Pvu*II αποτελεί η περιοχή Ser133-Lys137 που η μέχρι τώρα ερεύνα δεν της έχει αποδώσει κάποιο λειτουργικό ρόλο.



Εικόνα 38. Διάγραμμα Luzatti {Luzzati, 1952}. Κατανομή του παράγοντα R σε σχέση με την διακριτικότητα *{Biomol System Groningen, 1991}. Παρουσιάζονται επίσης οι θεωρητικές καμπύλες για 5 τιμές των Rms σφαλμάτων συντεταγμένων. Για την R.PvuII το rms σφάλμα στις συντεταγμένες μπορεί να εκτιμηθεί μεταξύ 0.2 και 0.3Å.*

Β 9. Ανάλυση του μοντέλου της R.*Pvu*II.

Β 9.1. Χαρακτηρισμός των στοιχείων της δευτεροταγούς δομής.

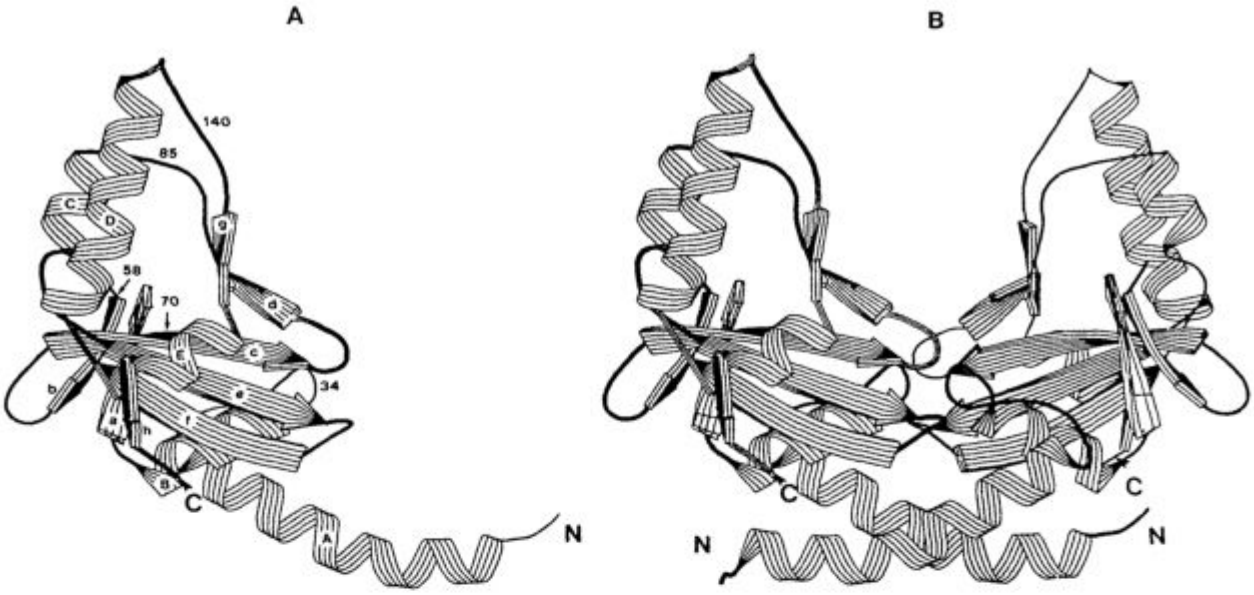
Ο χαρακτηρισμός των στοιχείων της δευτεροταγούς δομής του μορίου πραγματοποιήθηκε με το πρόγραμμα DSSP {Kabsch & Sander, 1983} του πακέτου CCP4 (πίνακας 20, εικόνα 50α). Το πρόγραμμα RIBBON {Priestle, 1988} χρησιμοποιήθηκε για την απεικόνιση της θέσης στον χώρο των στοιχείων της δευτεροταγούς δομής και το αποτέλεσμα παρουσιάζεται στην εικόνα 39. Στις έλικες του μορίου αποδόθηκαν τα κεφαλαία γράμματα A-E ενώ στους β-κλώνους τα μικρά γράμματα a-h και έτσι θα αναφέρονται στην συνέχεια του κειμένου. Οι β-κλώνοι βa, βb, βd, βg και βh επεκτείνονται σε περισσότερα κατάλοιπα από αυτά του πίνακα 20 αν χρησιμοποιηθούν ως κριτήριο μόνο οι φ και ψ γωνίες των καταλοίπων καθώς λόγω της παραμόρφωσης της κεντρικής β-πτυχωτής επιφάνειας μερικοί από τους προβλεπόμενους υδρογονικούς δεσμούς απουσιάζουν.

<u>Πίνακας 20. Αντιστοίχηση αλληλουχίας και δευτεροταγούς δομής.</u> Με τα κεφαλαία γράμματα A-E χαρακτηρίζονται οι α-έλικες του μορίου και με τα μικρά γράμματα a-h οι β-κλώνοι. Η Met1(*) απουσιάζει από την ώριμη πρωτείνη και τα κατάλοιπα Arg54-Glu55 λείπουν από το μοντέλο εξαιτίας της απουσίας ηλεκτρονιακής πυκνότητας στην περιοχή.

```
     *  AAAAAAAAAAAAAAAAAAAA          BBBBBBBBBBB  aa
  1  mSHPDLNKLLELWPHIQEYQDLALKHGINDIFQDNGGKLLQULLITGLTU

        **   bb     cccccccc     dd       CCCCCCC    ee
 51  LPGreGNDAUDNAGQEYELKSINIDLTKGFSTHHHMNPUIIAKYAQUPWI

     eeeee  ffffffff DDDDDDDDDDDDDDDDD          ggEEEEE
101  FAIYRGIAIEAIYALEPKDLEFYYDKWERKWYSDGHKDINNPKIPUKYUM

     EE hh
151  EHGTKIY
```

Β 9.2. Η δομή του μορίου στις τρείς διαστάσεις.

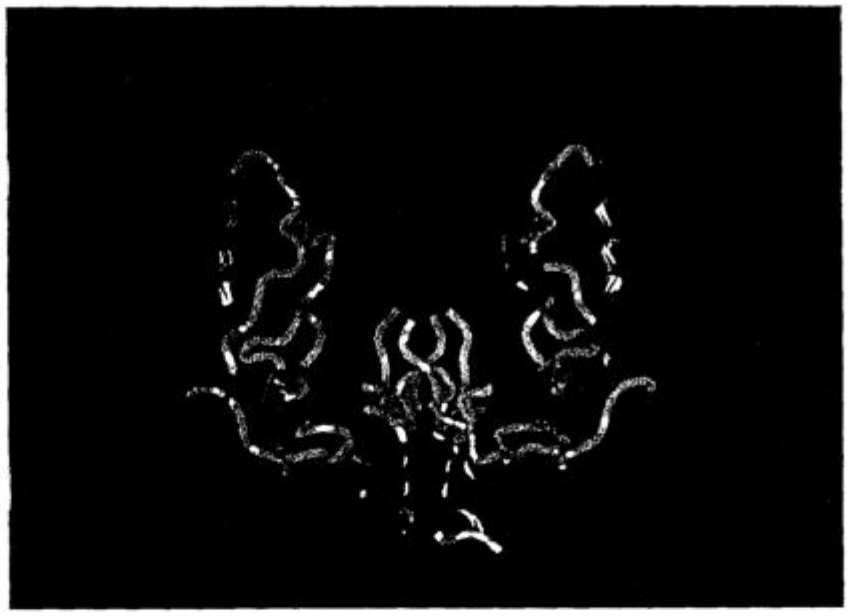
Στις εικόνες 39β και 40 γίνεται παράσταση του διμερούς της *Pvu*II ενδονουκλεάσης σε απλουστευμένη μορφή που περιλαμβάνει τα κύρια στοιχεία της δευτεροταγούς δομής της πρωτεΐνης. Το διμερές της πρωτεΐνης έχει μορφή U με συνολικές διαστάσεις 70x55x30Å. Ο τάξεως-2 άξονας αντιστοιχεί σε μία περιστροφή του κάθε μονομερούς 179°. Τα δύο μονομερή, μεγέθους 156



Εικόνα 39. Απεικόνιση σε μορφή κορδέλας της δομής της PvuII περιοριστικής ενδονουκλεάσης. *Α) Το μονομερές. Σημειώνεται η ονομασία των δευτεροταγών στοιχείων καθώς και τα αμίνο- και καρβόξυ-τελικά άκρα της πρωτεΐνης. Β) Το διμερές.*

αμινοξέων το καθένα, παρουσιάζουν πολύ μικρές διαφορές με μία rms απόσταση 0.4Å για τα άτομα της κύριας αλυσίδας (εξαιρώντας τα N-τελικά αμινοξέα Ser2-Pro4, Gly53 και Gly56 (σπάσιμο του μοντέλου) που έχουν την μεγαλύτερη απόκλιση). Η πρωτεΐνη έχει μικτή α/β μοριακή αρχιτεκτονική {Richardson, 1985} με τον πυρήνα της να σχηματίζεται από μία μικτή παράλληλη/αντιπαράλληλη β-πτυχωτή επιφάνεια αποτελούμενη από 6 β-κλώνους (βa, βb, βc, βe, βf, βh). Μία μικρότερη β-πτυχωτή επιφάνεια αποτελούμενη από τους αντιπαράλληλους κλώνους βd και βg σχηματίζει την εσωτερική επιφάνεια του, U-σχήματος, διμερούς. Τέλος η μία όψη της κεντρικής β-πτυχωτής επιφάνειας καταλαμβάνεται από τις έλικες αC, αD και αE ενώ η άλλη από τις αμινοτερματικές έλικες αA και αB. Η έλικα αA αποτελεί την βάση του μορίου και εμφανίζει μια κεντρική κάμψη 30° εξαιτίας της Pro14.

Κάθε μονομερές είναι οργανωμένο σε δύο δομικά ανεξάρτητες υποπεριοχές (domains) {Janin & Chothia, 1985}. Η υποπεριοχή A (Ser2-Asn35) που περιλαμβάνει την έλικα αA καθώς και τον βρόγχο (His26-Asn35) μεταξύ των έλικων αA και αB, σχηματίζει το μεγαλύτερο μέρος των αλληλεπιδράσεων διμερισμού και θα αναφέρεται στην συνέχεια του κειμένου ως υποπεριοχή διμερισμού. Η υποπεριοχή B περιλαμβάνει το υπόλοιπο της πρωτεΐνης. Οι ελάχιστες αλληλεπιδράσεις ανάμεσα στις δύο υποπεριοχές των μονομερών



Εικόνα 40. Απλουστευμένη παράσταση της δομής του διμερούς R.PvuII *Το μόριο περιεστραμένο κατά 90° σε σχέση με την εικόνα 39β γύρο από τον τάξεως-2 άξονα συμμετρίας. Κόκκινες οι α-έλικες και πράσινες οι β-πτυχωτές επιφάνειες του μορίου. Λόγω περιορισμών του προγράμματος "O" β-κλώνοι της κεντρικής β-πτυχωτής επιφάνειας με μικρό αριθμό καταλοίπων απουσιάζουν.*

συμβαίνουν ανάμεσα σε άτομα της κύριας αλυσίδας καταλοίπων της υποπεριοχής A και των καταλοίπων της πρώτης στροφής της έλικας αB (πίνακας 23). Στην υποπεριοχή αυτή ανήκει και η έλικα αB που συμμετέχει τόσο στις αλληλεπιδράσεις διμερισμού (πίνακας 22) όσο και σε αλληλεπιδράσεις με κατάλοιπα από τους β-κλώνους βe και βf λειτουργώντας έτσι ως γέφυρα ανάμεσα στην υποπεριοχή A του ενός μονομερούς και την υποπεριοχή B του δεύτερου.

Παρά την συμμετοχή μικρού τμήματος της πρωτεΐνης στις αλληλεπιδράσεις διμερισμού, η σταθερότητα του διμερούς επιτυγχάνεται μέσω ενός ιδιότυπου "αγκαλιάσματος" των μονομερών (εικόνα 41). Η έλικα αA τρέχει αντιπαράλληλα με την αντίστοιχη έλικα αA' του δεύτερου μονομερούς, σχηματίζοντας το κύριο μέρος των αλληλεπιδράσεων διμερισμού. Οι υπόλοιπες αλληλεπιδράσεις ανάμεσα στα δύο μονομερή πραγματοποιούνται από τον βρόγχο (LAB) που συνδέει τις έλικες αA και αB (His26-Asn35) και τμήματα της έλικας αB. Το πρίν την Pro14 τμήμα της έλικας αA κάθε μονομερούς μαζί με το μετά την Pro14 τμήμα του άλλου μονομερούς και την έλικα αB, σχηματίζουν ένα ψεύδο-δεμάτι τριών α-ελίκων με τις πλευρικές αλυσίδες των υδρόφοβων κατάλοιπων (κυρίως Ile, Leu εικόνα 44) να δείχνουν στο εσωτερικό του δεματιού, δημιουργώντας ένα ισχυρό υδρόφοβο

πυρήνα (πίνακας 22). Οι υπόλοιπες αλληλεπιδράσεις είναι υδρογονικοί δεσμοί {Baker & Hubbard, 1986} και δεσμοί αλάτων μεταξύ ατόμων των πλευρικών αλυσίδων αλλά και της κύριας αλυσίδας στην περιοχή του βρόγχου LAB κοντά στον διπλό άξονα συμμετρίας (πίνακας 21).

Ανάμεσα στις α-έλικες αC και αD και στην μικρή β-πτυχωτή επιφάνεια που αποτελείται από τους β-κλώνους βd και βg, δημιουργείται κενός χώρος στον οποίο πακετάρονται δακτύλιοι αρωματικών αμινοξέων Trp, Tyr και Phe (εικόνα 43) συνεισφέροντας στο σφικτό πακετάρισμα της υποπεριοχής B.

Πίνακας 21. Πιθανοί υδρογονικοί δεσμοί μεταξύ των μονομερών της R.*Pvu*II. Στα σχόλια αναφέρονται τα στοιχεία της δευτεροταγούς δομής στην οποία ανήκουν τα κατάλοιπα που άτομα τους μετέχουν στην δημιουργία του δεσμού. LAB ο βρόγχος His26-Asn35 που συνδέει τις α-έλικες αA και αB ενώ Lef ο βρόγχος Gly106-Ile107 μεταξύ των β-κλώνων e και f.

Μονομερές A Κατάλοιπο	Ατομο	Μονομερές B Κατάλοιπο	Ατομο	Απόσταση (Å)	Σχόλια
Asp 5	$O^{\delta 1}$	Lys 25	N^{ζ}	3.23	αA-αA
Asp 5	$O^{\delta 2}$	His 26	$N^{\varepsilon 2}$	2.93	αA-LAB
Lys 8	N^{ζ}	Lys 25	N^{ζ}	2.87	αA-αA†
Glu 18	$O^{\varepsilon 1}$	Lys 8	N^{ζ}	2.92	αA-αA
His 26	$N^{\varepsilon 2}$	Asp 5	$O^{\delta 1}$	2.62	LAB-αA
Asp 30	$O^{\delta 1}$	Lys 38	N^{ζ}	2.81	LAB-αB
Phe 32	O	Asn 35	N	2.99	LAB-LAB
Gln 33	$O^{\varepsilon 1}$	Asn 35	$N^{\delta 2}$	3.15	LAB-LAB
Asn 35	N	Phe 32	O	3.00	LAB-LAB
Lys 38	N^{ζ}	Asp 30	Oδ2	3.37	αB-LAB
Leu 43	O	His 3	$N^{\delta 1}$	2.67	αB-N
Leu 44	O	His 3	N	3.01	αB-N
Ile 45	O	Ser 2	O^{γ}	2.70	αB-N
Ile 45	O	Ser 2	N	2.56	αB-N

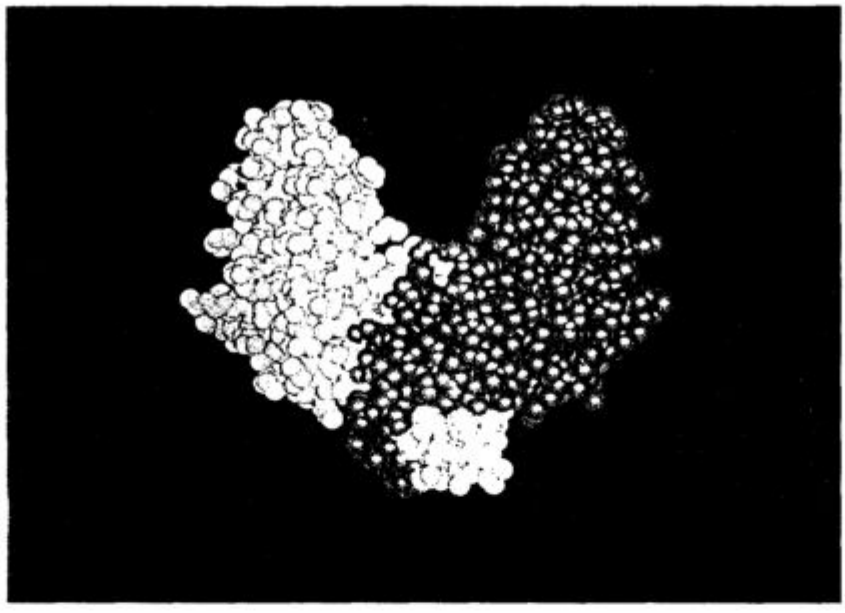
† Πρόκειται πιθανώς για στενή επαφή και όχι υδρογονικό δεσμό.

Πίνακας 22. Υδρόφοβες αλληλεπιδράσεις μεταξύ των μονομερών της R.*Pvu*II. Στον πίνακα παρουσιάζονται τα κατάλοιπα, που μπορούν να μετάσχουν σε υδρόφοβες αλληλεπιδράσεις, των οποίων άτομα C της πλευρικής αλυσίδας έρχονται σε απόσταση μικρότερη από 5Å με άτομα αντίστοιχων καταλοίπων του δεύτερου μονομερούς. Στα σχόλια σημειώνονται οι δευτεροταγείς δομές στις οποίες ανήκουν τα κατάλοιπα. LAB ο βρόγχος His26-Asn35 που συνδέει τις α-έλικες αA και αB ενώ Lef ο βρόγχος Gly106-Ile107 μεταξύ των β-κλώνων e και f.

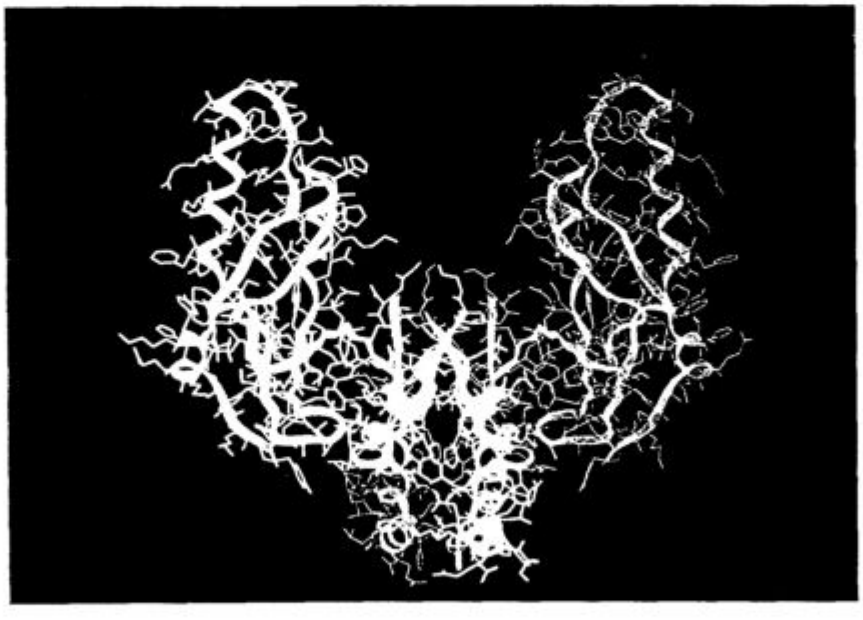
Μονομερές A Κατάλοιπο	Ατομο	Μονομερές B Κατάλοιπο	Ατομο	Σχόλιο
Leu 6	$C^{\alpha},C^{\delta 1}$	Leu 44	C^{β}	αA-αB
Leu 6	$C^{\delta 1}$	Ile 45	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\gamma 2}$	αA-αB
Lys 8	$C^{\beta},C^{\gamma},C^{\delta},C^{\varepsilon}$	Leu 22	$C^{\gamma},C^{\delta 1},C^{\delta 2}$	αA-αA
Leu 9	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta 1},C^{\delta 2}$	Leu 44	$C^{\beta},C^{\delta 1}$	αA-αB
Leu 9	$C^{\delta 2}$	Ile 45	$C^{\varepsilon 1},C^{\gamma 1}$	αA-αB
Leu 12	$C^{\delta 1}$	Leu 22	$C^{\delta 2}$	αA-αA
Ile 16	$C^{\beta},C^{\gamma 2}$	Ile 107	$C^{\delta 1}$	αA-Lef
Ile 16	$C^{\delta 1}$	Ile 16	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\delta 1}$	αA-αA
Leu 22	$C^{\delta 2}$	Leu 9	$C^{\alpha},C^{\delta 1}$	αA-αA
Leu 22	$C^{\delta 1},C^{\delta 2}$	Lys 8	C^{β},C^{γ}	αA-αA
Ile 31	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\delta 1}$	Phe 32	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta 1},C^{\varepsilon 1}$	LAB-LAB
Phe 32	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta 1},C^{\varepsilon 1}$	Ile 31	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\varepsilon 1}$	LAB-LAB
Phe 32	$C^{\alpha},C^{\beta},C^{\delta 1},C^{\varepsilon 1}$	Phe 32	$C^{\alpha},C^{\beta},C^{\gamma},C^{\varepsilon 1},C^{\varepsilon 1}$	LAB-LAB
Phe 32	$C^{\beta},C^{\gamma},C^{\delta 1},C^{\delta 2},C^{\varepsilon 1},C^{\varepsilon 2},C^{\zeta}$	Gly 37	C^{α}	LAB- αB
Phe 32	$C^{\gamma},C^{\delta 2},C^{\varepsilon 2}$	Lys 38	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta},C^{\varepsilon}$	LAB-αB
Phe 32	$C^{\gamma},C^{\delta 2},C^{\varepsilon 2},C^{\zeta}$	Ile 107	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\gamma 2},C^{\delta 1}$	LAB-Lef
Gly 37	C^{α}	Phe 32	$C^{\beta},C^{\gamma},C^{\varepsilon 1},C^{\varepsilon 2},C^{\varepsilon 1},C^{\varepsilon 2},C^{\zeta}$	αB-LAB
Lys 38	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta},C^{\varepsilon}$	Phe 32	$C^{\gamma},C^{\delta 2},C^{\varepsilon 2}$	αB-LAB
Leu 44	$C^{\beta},C^{\gamma},C^{\delta 1}$	Leu 9	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta 1},$	αB-αA
Leu 44	$C^{\beta},C^{\delta 1}$	Leu 6	$C^{\alpha},C^{\beta},C^{\gamma},C^{\delta 1}$	αB-αA
Ile 45	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\delta 1}$	Leu 6	$C^{\delta 1}$	αB-αA
Ile 45	$C^{\delta 1},C^{\gamma 1}$	Leu 9	$C^{\delta 2}$	αB-αA
Ile 107	$C^{\alpha},C^{\beta},C^{\gamma 1},C^{\gamma 2},C^{\delta 1}$	Phe 32	$C^{\gamma},C^{\delta 2},C^{\varepsilon 2},C^{\zeta}$	Lef-LAB
Ile 107	$C^{\delta 1}$	Ile 16	$C^{\beta},C^{\gamma 2}$	Lef-αA

Πίνακας 23. Πιθανοί υδρογονικοί δεσμοί μεταξύ των δύο υποπεριοχών A (Ser2-Asn35) και B (Gly36- Tyr157) της πρωτεΐνης.

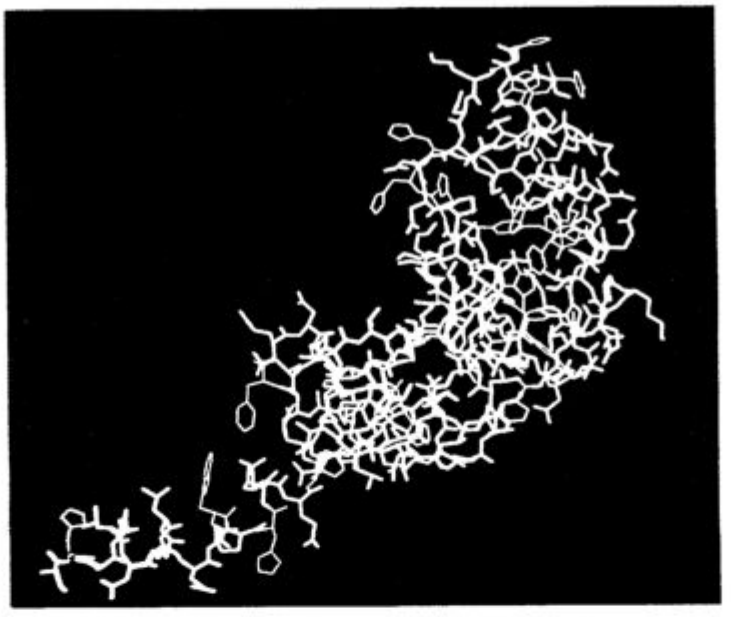
Υποπεριοχή A Κατάλοιπο	Ατομο	Υποπεριοχή B Κατάλοιπο	Ατομο	Απόσταση (Å)	Σχόλια
Tyr 19	OH	Gly 37	O	2.66	αA-αB
Ile 31	O	Gly 37	N	2.96	LAB-αB
Asn 35	O	Leu 39	N	2.96	LAB-αB
Asn 35	O	Lys 38	N	2.69	LAB-αB



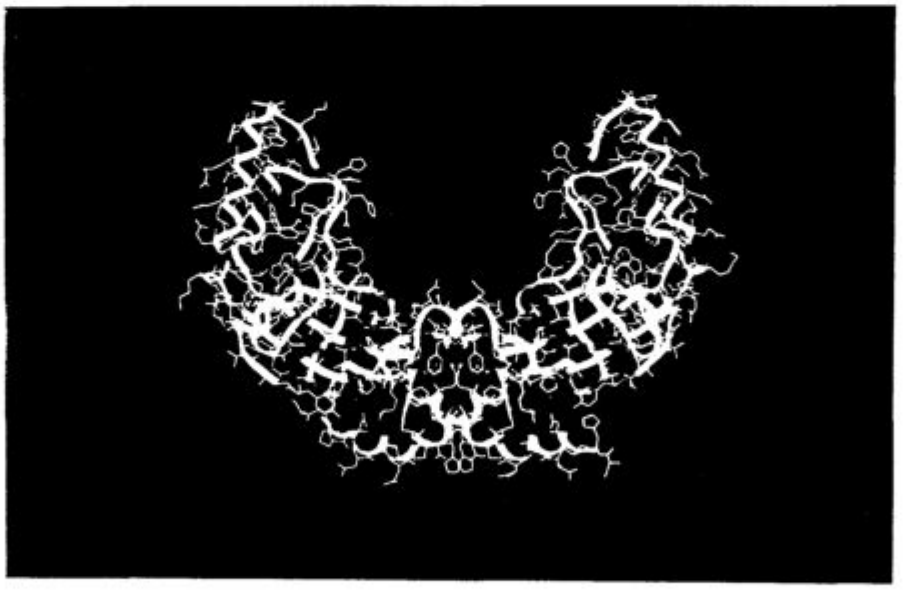
Εικόνα 41. Επιφάνεια αλληλεπίδρασης ανάμεσα στα μονομερή της R.PvuII. *Σε χωροπληρωτικό μοντέλο του διμερούς της R.PvuII με διαφορετικό χρώμα διακρίνονται τα δύο μονομερή.*



Εικόνα 42. Το διμερές της R.PvuII με εναποτιθεμένο μοντέλο κορδέλας. *Στο κέντρο διακρίνεται η περιοχή αλληλεπίδρασης των δύο μονομερών.*



Εικόνα 43. Αρωματικά αμινοξέα στο μόριο της R.PvuII. *Στο κέντρο της υποπεριοχής Β ο "αρωματικός" πυρήνας του μορίου.*



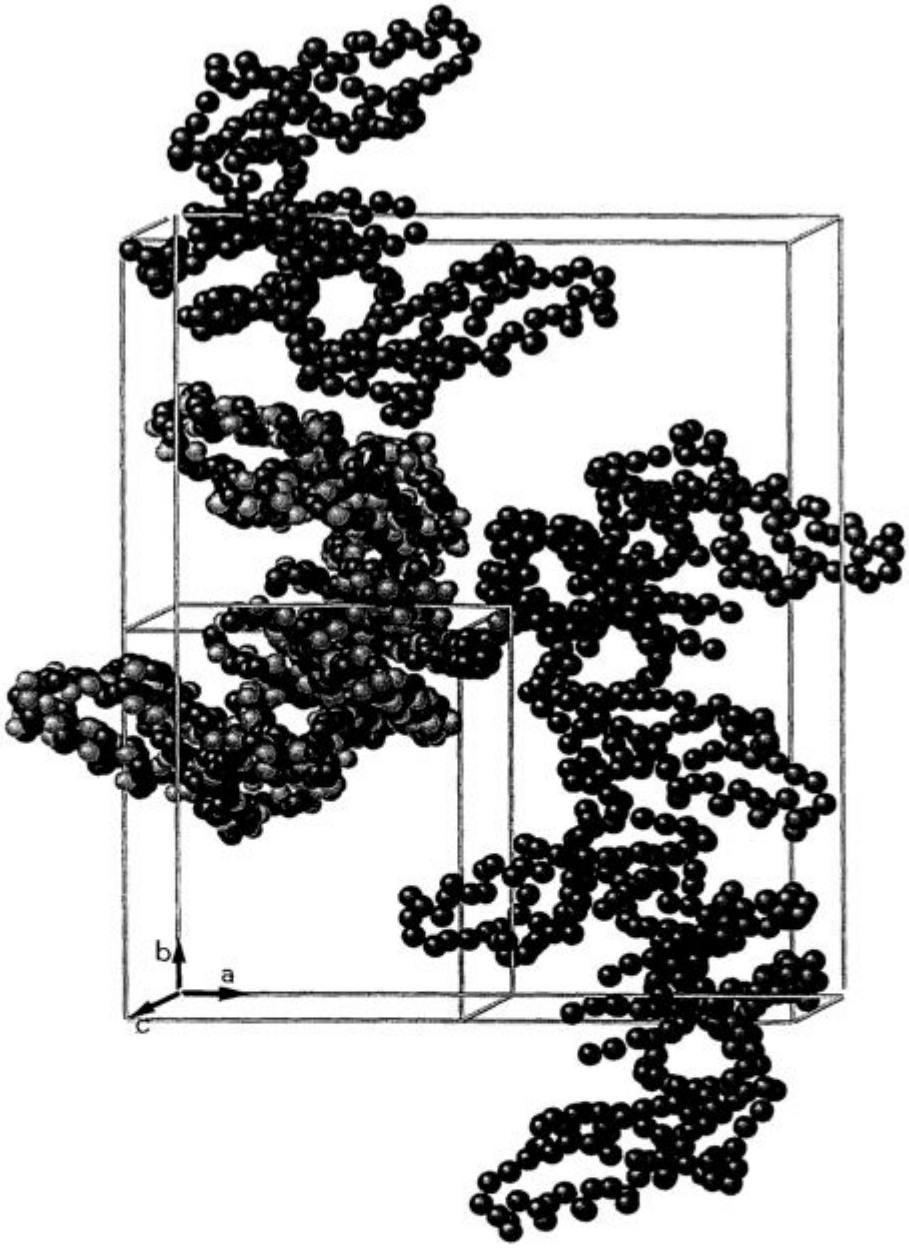
Εικόνα 44. Κατανομή των καταλοίπων Leu (μπλέ) και Ile (κόκκινο) στο διμερές μόριο της R.PvuII. *Γίνεται εμφανής η αμφιπαθηκότητα της έλικας αΑ και η ευρύτερη περιοχή του υδρόφοβου πυρήνα διμερισμού του μορίου.*

Β 9.3. Πακετάρισμα των μορίων στην κρυσταλλική δομή.

Οπως έχει αναφερθεί προηγούμενα η στοιχειώδης κυψελίδα που συνθέτει τον κρύσταλλο της R.*Pvu*II περιλαμβάνει συνολικά 4 διμερή μόρια τα οποία σχετίζονται μεταξύ τους με τις κρυσταλλογραφικές συμμετρίες που ορίζονται για την $P2_12_12$ χωροομάδα. Τα δύο μονομερή καθενός από τα διμερή εμφανίζουν

Πίνακας 24. Κρυσταλλικές επαφές. Στον πίνακα σημειώνονται τα άτομα τα οποία μετέχουν σε υδρογονικούς δεσμούς και γέφυρες αλάτων μεταξύ διαφορετικών διμερών συνδεόμενων με τις κρυσταλλογραφικές συμμετρίες. Στην στήλη της συμμετρίας αναφέρεται ο μετασχηματισμός που συνδέει τα μόρια που μετέχουν στον δεσμό Υδρογονικοί δεσμοί με μειωμένη πιθανότητα σημειώνονται με * στην στήλη των σχολίων. Οι υπολογισμοί των αποστάσεων έγιναν με το πρόγραμμα Contact {CCP4, 1979} για αποστάσεις ατόμου - ατόμου 2.0-3.5Å.

Μόριο Α Κατάλοιπο	Ατομο	Συμμετρία	Μόριο Β Κατάλοιπο	Ατομο	Απόσταση (Å)	Σχόλια
Ser A2	O^{γ}	1-x,1-y,z	Ile B45	O	2.71	
Ser A2	O	1-x,1-y,z	Ser A2	N,O	2.84, 2.78	
Ser A2	N	1-x,1-y,z	Ser A2	O	2.84	
Asn A62	O	x,y,z-1	W 526	OH2	3.09	
Gly A64	O	x,y,z-1	W 563	OH2	3.45	*
Glu A110	O	1/2-x,y-1/2,1-z	Arg B129	NH^2	2.70	
Tyr A132	O	x,1-y,z-1	Gln B196	$N^{\varepsilon 2}$	2.60	
Met A150	O	1/2-x,y-1/2,1-z	Lys B118	N^{ζ}	3.01	
Glu A151	$O^{\varepsilon 2}$	1/2-x,y-1/2,1-z	Asp B119	$O^{\delta 1}$	2.86	Στενή επαφή;
Lys A155	N^{ζ}	1/2-x,y-1/2,1-z	Asp B125	$O^{\delta 2}$	2.81	
Tyr A157	O^{T}	1/2-x,y-1/2,1-z	Arg B129	NH^1	2.69	
Tyr A132	O	-x,1-y,1-z	Gln B96	$N^{\varepsilon 2}$	2.60	
Ile B45	O	1-x,1-y,z	Ser A2	O^{γ}	2.71	
Ile B45	O	1-x,1-y,z	W568	OH2	3.35	*
Glu B66	$O^{\varepsilon 1}$	x,y,1+z	W543	OH2	2.75	
Gln B96	$N^{\varepsilon 2}$	-x,1-y,1+z	Tyr A132	O	2.60	
Pro B117	N	-x+1/2,y+1/2,1-z	W495	OH2	3.31	*
Lys B118	N^{ζ}	-x+1/2,y+1/2,1-z	Met 150	O	3.01	
Lys B118	N	-x+1/2,y+1/2,1-z	W495	OH2	2.75	
Asp B119	$O^{\delta 1}$	-x+1/2,y+1/2,1-z	Glu 151	$O^{\varepsilon 2}$	2.86	Στενή επαφή;
Phe B122	O	-x+1/2,y+1/2,1-z	W505	OH2	2.68	
Asp B125	$O^{\delta 2}$	-x+1/2,y+1/2,1-z	Lys 155	N^{ζ}	2.81	
Lys B126	O	-x+1/2,y+1/2,1-z	W502	OH2	3.01	
Arg B129	NH^1	-x+1/2,y+1/2,1-z	W476	OH2	2.81	
Arg B129	NH^2	-x+1/2,y+1/2,1-z	Glu 110	O	2.70	
Lys B130	N	-x+1/2,y+1/2,1-z	W502	OH2	3.23	
Tyr B132	O	-x+1/2,y+1/2,1-z	W501	OH2	3.22	
Glu B151	$O^{\varepsilon 1}$	-x+1/2,y+1/2,1-z	W553	OH2	2.66	
Lys B155	N^{ζ}	1-x,1-y,z	W567	OH2	3.12	
Tyr B157	O^{T}	1-x,1-y,z	W497	OH2	3.49	*



Εικόνα 45. Παράσταση του πακεταρίσματος των μορίων της R.PvuII στο κρυσταλλικό πλέγμα. *Σε μορφή χωροπληρωτικού μοντέλου εμφανίζονται τα άτομα τις κύριας αλυσίδας για ένα διμερές μόριο ενώ για τα συμμετρικά του μόνο τα C^{α} άτομα. Το εξωτερικό παραλληλεπίπεδο παριστάνει μία στοιχειώδη κυψελίδα ενώ το εσωτερικό την ασύμμετρη μονάδα των κρυστάλλων Η παράσταση έγινε με το πρόγραμμα Sybyl {Tripos, 1992}.*

μεταξύ τους μικρές διαφορές που μπορούν να αποδοθούν στο διαφορετικό περιβάλλον στο οποίο βρίσκεται καθένα από αυτά. Στην εικόνα 45 εμφανίζεται σε απλουστευμένη μορφή η διευθέτηση των μορίων που περιλαμβάνονται σε μία

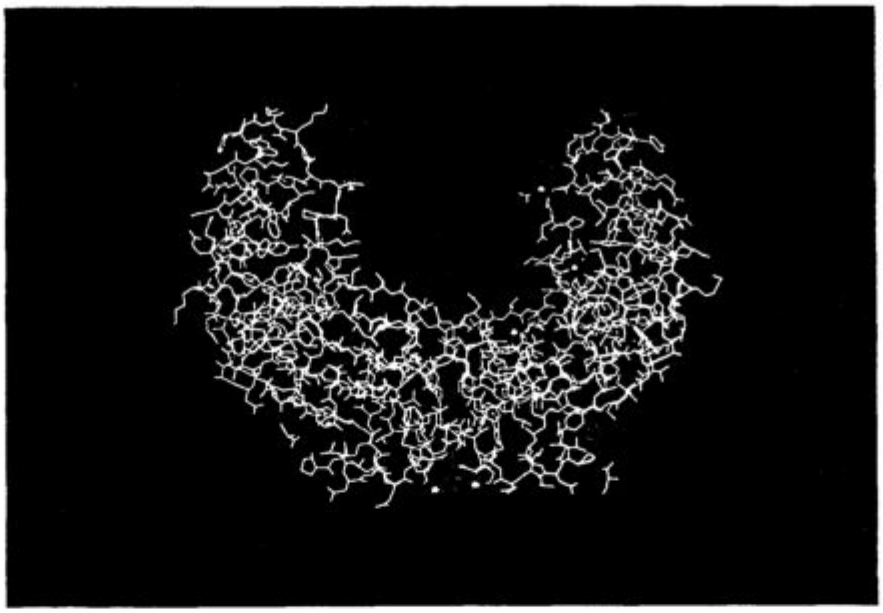
στοιχειώδη κυψελίδα. Η διαφοροποίηση του περιβάλλοντος κάθε μονομερούς αντανακλάται στις αλληλεπιδράσεις του κάθε μονομερούς με άτομα γειτονικών διμερών (πίνακας 24).

Β 9.4. Θέσεις της πρόσδεσης των βαρειών ατόμων στην πρωτεΐνη.

Εχοντας πλέον το τελικό μοντέλο του μορίου της R.*Pvu*II, ανατρέξαμε στις θέσεις πρόσδεσης των βαρειών ατόμων που χρησιμοποιήθηκαν στον προσδιορισμό των φάσεων των παραγόντων δομής *F* της πρωτεΐνης, προκειμένου να ταυτοποιήσουμε τα κατάλοιπα της πρωτεΐνης που μετέχουν στην δημιουργία των θέσεων πρόσδεσης τους. Στον πίνακα 25 παρουσιάζονται τα άτομα του μοντέλου της R.*Pvu*II που απέχουν απόσταση μικρότερη των 4Å από τις θέσεις πρόσδεσης των βαρειών άτομων και αποτελούν πιθανούς παράγοντες πρόσδεσης τους. Οι αποστάσεις που σημειώνονται στον πίνακα αυτό είναι βασισμένες στο μοντέλο της φυσικής πρωτεΐνης έτσι είναι πιθανό ότι μικρές διαφοροποιήσεις στις αποστάσεις αυτές μπορεί να συμβαίνουν όταν γίνεται η πρόσδεση του βαρέως ατόμου.

Από τις 9 κύριες και δευτερεύουσες θέσεις πρόσδεσης ατόμων αργύρου ανά διμερές της πρωτεΐνης, οι 8 συνδέονται ανά δύο με τον άξονα της μή κρυσταλλογραφικής συμμετρίας. Ολες οι θέσεις πρόσδεσης σχηματίζονται με την συμμετοχή των ιμιδαζολικών πλευρικών ομάδων καταλοίπων His. Κάθε μονομερές της R.*Pvu*II περιλαμβάνει 8 κατάλοιπα His στην αλληλουχία του από τις οποίες οι 7 (His3, His15, His26, His83, His84, His85, His152) συμμετέχουν σε θέσεις δέσμευσης για το άτομο του Ag. Η κύρια θέση πρόσδεσης του Ag είναι η μόνη στην οποία το άτομο αργύρου είναι συνδεδεμένο με το ένα μόνο μονομερές. Αυτό εξηγείται από το γεγονός ότι η θέση αποτελεί ταυτόχρονα κρυσταλλική επαφή μεταξύ διμερών τα οποία σχετίζονται με την κρυσταλλογραφική συμμετρία. Πλησιέστερα άτομα στον Ag1 είναι το S^{δ} άτομο της Met150 του μονομερούς A1 και η -NH ομάδα της His152 του μονομερούς B2 ενός διμερούς που προκύπτει από την εφαρμογή της κρυσταλλογραφικής συμμετρίας. Ο μή σχηματισμός της αντίστοιχης θέσης στο δεύτερο μονομερές οφείλεται στο διαφορετικό κρυσταλλικό περιβάλλον της Met150 όπως φαίνεται στον πίνακα 24. Η συμβολή του Ag σε κρυσταλλική επαφή αποτελεί πιθανότατα και εξήγηση του φαινομένου : οι κρύσταλλοι που αναπτύσονται σε πειράματα συγκρυστάλλωσης $AgNO_3$ και πρωτεΐνης, να είναι καλύτερης ποιότητας και μεγαλύτερου μεγέθους από αυτούς της φυσικής πρωτεΐνης κάτω από ίδιες συνθήκες {δεδομένα που δεν παρουσιάζονται}.

Τόσο οι κύριες όσο και δευτερεύουσες θέσεις πρόσδεσης για τα δύο παράγωγα λανθανίδων (Yb,Pr) που χρησιμοποιήθηκαν ταυτίζονται. Η δευτερεύουσες θέσεις πρόσδεσης είναι γειτονικές προς τις κύριες σχηματίζοντας συνάθροιση (clustering) βαρειών ατόμων στην ίδια περιοχή της πρωτεΐνης και έχει παρατηρηθεί και σε άλλες περιπτώσεις παραγώγων λανθανίδων {Blundell & Johnson, 1976}. Στην δημιουργία της θέσεως πρόσδεσης, έχουν συμμετοχή οι πλευρικές καρβοξυλικές ομάδες των αμινοξέων Asp58 και Glu68. Τα δύο αυτά



Εικόνα 46 . Οι κύριες θέσεις πρόσδεσης των βαρειών ατόμων στο διμερές μόριο της R.PvuII. *Hg -Thiomersal (μπλέ σφαίρες), Pr (κόκκινες σφαίρες), και Ag (πράσινες σφαίρες).*

αμινοξέα συμμετέχουν στο σχηματισμό της ενεργού περιοχής του ενζύμου και αποτελούν πιθανούς παράγοντες πρόσδεσης Mg^{2+} (βλ. κεφάλαιο B10.1.), {Cheng et al., 1994}.

Η απουσία καταλοίπων Cys από την πρωτοταγή αλληλουχία της R.*Pvu*II πρωτεΐνης η οποία αποτελεί τον κατεξοχήν παράγοντα πρόσδεσης υδραργυρικών ενώσεων, δεν εμπόδισε την παραγώγιση των κρυστάλλων της R.*Pvu*II με την ένωση του υδραργύρου, Thiomersal (εικόνα 16). Τον ρόλο της Cys έπαιξε η Ser71 ενώ πιθανά συμμετοχή στην δημιουργία της θέσης πρόσδεσης έχει και η Tyr104. Το μεγάλο μέγεθος του μορίου Thiomersal καθιστά πιθανώς, την ύπαρξη κοιλοτήτων στο μόριο τον κύριο παράγοντα για την ειδικότητα της παραγώγισης.

Πίνακας 25. Κατάλοιπα με πιθανή συμμετοχή στην πρόσδεση των βαρίων ατόμων. Στον πίνακα σημειώνονται τα άτομα της πρωτείνης που απέχουν λιγότερο απο 5Å από θέσεις πρόσδεσης βαρειών ατόμων Οπου περισσότερα από ένα άτομα ενός καταλοίπου, είναι σε θέση πιθανής αλληλεπίδρασης, αυτά εμφανίζονται σε παρένθεση. Τα δύο μονομερή της R.*Pvu*II διακρίνονται ως A και B στην ονοματολογία των καταλοίπων. Σε αγκύλες τα δευτεροταγή στοιχεία στα οποία ανήκουν τα κατάλοιπα με συμμετοχή στις θέσεις πρόσδεσης, ως LXX συμβολίζεται ένας βρόγχος όπου XX είναι τα δευτεροταγή στοιχεία τα οποία συνδέει.

Βαρύ άτομο	Κατάλοιπα	Πρωτεινικό Ατομο	Απόσταση από το βαρύ άτομο (Å)	Σχόλια
Pr1	Asp58B [Lab], Glu68B [βc]	($O^{\delta1}$, $O^{\delta2}$), ($O^{\varepsilon1}$, $O^{\varepsilon2}$)	3.04,1.88 3.51,3.58	Συμμετέχουν στον σχηματισμό της καταλυτικής περιοχής. (ΠρόσδεσηMg^{2+} ;)
Pr2	Asp58A [Lab], Glu68A [βc]	($O^{\delta1}$, $O^{\delta2}$), ($O^{\varepsilon1}$, $O^{\varepsilon2}$)	2.18,2.54 1.82,2.85	»
Hg1	Ser71B [βc] Tyr104B [βe]	O^{γ}, OH	3.91,3.80	
Hg2	Ser71A [βc], Tyr104A [βe]	O^{γ}, OH	3.53,4.78	
Ag1	His152B [αE], Tyr148B [αE], Asp119B [αD], Met150 [αE]**	($N^{\delta1}$,$N^{\varepsilon2}$), OH,$O^{\delta1}$ S^{δ}	(2.34,3.56) 3.74,4.51 2.51	
Ag2	His85A [LdC]	($N^{\delta1}$,$N^{\varepsilon2}$)	(0.91,2.96)	Περιοχή αναγνώρισης του DNA.
Ag3	His15B [αA], Glu18B [αA], Glu11A [αA]	$N^{\varepsilon2}$, $O^{\varepsilon2}$	2.48 3.67,3.43	
Ag4	His84B [LdC], His85B [LdC]	O,N	3.93,3.21	Περιοχή αναγνώρισης του DNA.
Ag5	His26A [LAB], His3B [N], Asp5B [αA]	($N^{\delta1}$,$N^{\varepsilon2}$), ($N^{\delta1}$,$N^{\varepsilon2}$), $O^{\delta1}$	(3.14,3.19) (2.15,2.96), 3.41	
Ag6	His15A [αA], His15B [αA]	$N^{\varepsilon2}$,$N^{\delta1}$	3.06,3.40	
Ag7	His3A [N], Asp5A [αA], His26B [LAB]	$N^{\varepsilon2}$,$O^{\delta2}$, ($N^{\delta1}$,$N^{\varepsilon2}$)	4.07,3.88 (3.04,3.09)	
Ag8	His85A [LdC]	($N^{\delta1}$,$N^{\varepsilon2}$)	(1.90,3.53)	
Ag9	His85B [Ldc], His84B [Ldc], His83B [Ldc]	($N^{\delta1}$,$N^{\varepsilon2}$),N $N^{\delta1}$	(3.76,4.34),4.72 4.46	Περιοχή αναγνώρισης του DNA.

** Το κατάλοιπο Met150 ανήκει σε κρυσταλλογραφικά σχετιζόμενο συμμετρικό μόριο.

Β 10. Συγκριτική μελέτη της δομής της R.*Pvu*II.

Β 10.1. Συσχέτιση των δομών της R.*Pvu*II και R.*Eco*RV.

Μια πρώτη παρατήρηση του μοντέλου της R.*Pvu*II υπέδειξε μία ομοιότητα της γενικής μορφής της πρωτεΐνης με αυτήν της R.*Eco*RV (εικόνα 47). Αυτό αποτέλεσε έκπληξη καθώς οι δύο αμινοξικές αλληλουχίες δεν εμφάνιζαν καμμία προφανή ομοιότητα (πίνακας 26) σε συγκρίσεις που είχαν προηγηθεί. Επιπλέον, με βάση τις δομικές εργασίες πάνω στα περιοριστικά ένζυμα *Eco*RI {Kim et al., 1990} και *Eco*RV {Winkler et al., 1993}, δεν υπήρχαν αναφορές για ομοιότητες στο συνολικό δίπλωμα των ενδονουκλεασών {Anderson, 1993}.

Προκειμένου να αποδώσουμε την ομοιότητα αυτή σε συγκεκριμένα στοιχεία των δύο δομών οδηγηθήκαμε στο να πραγματοποιήσουμε μια λεπτομερέστερη σύγκριση τους, σε επίπεδο δευτεροταγούς και τεταρτοταγούς δομής. Για τον σκοπό αυτό χρησιμοποιήθηκαν κυρίως οι δυνατότητες του προγράμματος "O" {Jones & Kjeldgaard 1992}, {Jones, Zou and Cowan, 1991}, για υπέρθεση μορίων ύστερα από τρισδιάστατη στοίχιση, βασισμένη στην ελαχιστοποίηση της rms απόστασης των "τοπολογικά ισοδύναμων C^{α} ατόμων" των δύο μορίων, με την μέθοδο των ελαχίστων τετραγώνων {Matthews & Rossman, 1985}. Για τις συγκρίσεις που περιγράφονται χρησιμοποιήθηκαν οι συντεταγμένες των μοντέλων της R.*Eco*RV που είναι κατατεθιμένες στην Brookhaven βάση δεδομένων με τους κωδικούς 1RVE (ελεύθερη πρωτεΐνη), 2RVE (σύμπλοκο με DNA μή συγγενούς αλληλουχίας) και 4RVE (σύμπλοκο με DNA

Πίνακας 26. Στοίχιση των αλληλουχιών των R.EcoRV και R.PvuII. *Η στοίχιση πραγματοποιήθηκε με το πρόγραμμα BESTFIT (Devereux, Haeberli and Smithies, 1984). Η στοίχιση δίνει ποσοστό ταυτότητας αμινοξέων 20.53% ενώ ο δείκτης ποιότητας Q=54.7 συγκρίσιμος με τον μέσο (από 10 στοιχίσεις) δείκτη ύστερα από τυχαιοποίηση της αλληλουχίας Q_r=51.0 +/- 2.8.*

```
                .         .         .         .         .
EcoRV  47 FSRPIINKIAEKHGYIVEEPKQQNHYPDFTLYKPSEPNKKIAIDIKTTYTN 97
          :|:| :||: |  ..| |        |.|:.|            :|...: :
PvuII   1 MSHPDLNKLLELWPHIQE.......YQDLAL..........KHGINDIFQD 34
                  .         .         .         .         .
       98 KENEKIK.FTLGGYTSFIRNNTKNIVYPFDQYIAHWIIGYVYTRVATRKSS 147
          .:.. :. : :.|.| :  .:..: |                  |... .
       35 NGGKLLQVLLITGLTVLPGREGNDAV................DNAGQEYE 68
                  .         .         .         .         .
      148 LKTYNINELNEIPKPYKGVKVFLQDKWVIAGDLAGSGNTTNIGSIHAHYKD 198
          ||..||:  .:.....  .|:..: :             .:..| | |::
       69 LKSINIDLTKGFSTHHHMNPVIIAKYR.............QVPWIFAIYRG 106
                  .         .         .         .         .
      199 FVEGKGIFDSEDEFLDYW...RNYERTSQLRNDKYNNISEYRNWIYRGRK 245
           :. .:|:  |.. |:::   :: .: |: :.|  |   . :  : :| |
      107 .IAIEAIYRLEPKDLEFYYDKWERKWYSDGHKDINNPKIPVKYVMEHGTK 155
```

συγγενούς αλληλουχίας).

Η τρισδιάστατη στοίχιση των δύο ενζύμων (πίνακες 27, 28, 29) αποκάλυψε την παρουσία ενός συνόλου στοιχείων δευτεροταγούς δομής που εμφανίζονται στις δύο δομές σε τοπολογικά ισοδύναμες θέσεις στον χώρο (εικόνες 48, 49, 50, 51, 52). Στην R.*Pvu*II το σύνολο αυτό περιλαμβάνει τις α-έλικες αB, αD και αE και τους β-κλώνους βa, βb, βc, βd, βe, βf, βg. Τα δευτεροταγή αυτά στοιχεία στην R.*Eco*RV είναι τοπολογικά ισοδύναμα με τις α-έλικες αB, αD, αE και τους β-κλώνους βc, βd, βe, βf, βg, βh, βj. Στην R.*Pvu*II η έλικα αB είναι ισοδύναμη με την N-τελική περιοχή της αB έλικας της R.*Eco*RV. Από τα προηγούμενα φαίνεται ότι η υποπεριοχή B της R.*Pvu*II σε μεγάλη έκταση σχετίζεται τοπολογικά με την περιοχή πρόσδεσης του DNA της R.*Eco*RV. Αντίθετα οι περιοχές διμερισμού των δύο ενζύμων δεν περιλαμβάνουν τοπολογικά ισοδύναμα δευτεροταγή στοιχεία.

Πίνακας 27. Υπέρθεση R.PvuII και R.EcoRV (rms αποστάσεις των C^{α}). *Η τρισδιάστατη στοίχιση (υπολογισμός και εφαρμογή του πίνακα τελεστών μετασχηματισμού) των δύο μορίων πραγματοποιήθηκε αρχικά με το υποπρόγραμμα LSQ_EXPLICIT του προγράμματος "O" για τους β-κλώνους βc και βf της R.PvuII και αντίστοιχα βe και βh για την R.EcoRV. Για κάθε περίπτωση ο πίνακας μετασχηματισμού βελτιστοποιήθηκε με το υποπρόγραμμα LSQ_IMPROVE έτσι ώστε να μεγιστοποιηθεί ο αριθμός των C^{α} ατόμων με συμμετοχή στην στοίχιση. Σε παρένθεση είναι ο αριθμός των C^{α} με συμμετοχή στην στοίχιση ύστερα από βελτιστοποίηση του πίνακα μετασχηματισμού. Τα κατάλοιπα που συμμετείχαν στον προσδιορισμό του πίνακα μετασχηματισμού και στον υπολογισμό της rms απόστασης, ύστερα από την βελτιστοποίηση, παρουσιάζονται στους πίνακες 28 και 29.*

Περιοχή καταλοίπων	rms διαφορές των C^{α} (Å)	Σχόλιο
66-73, 108-115 *Pvu*II 86-93, 163-170 *Eco*RV	5.789	Υπέρθεση διμερών.
Βελτιστοποίηση (25 C^{α} άτομα, Πίνακας 28)[1]	2.550	Υπέρθεση διμερών.
66-73, 108-115 *Pvu*II 86-93, 163-170 *Eco*RV	4.849	Υπέρθεση μονομερών
Βελτιστοποίηση (59 C^{α} άτομα, Πίνακας 29)	2.083	Υπέρθεση μονομερών

[1] Συμμετοχή στην στοίχιση αυτή είχαν 25 C^{α} άτομα από κάθε μονομερές.

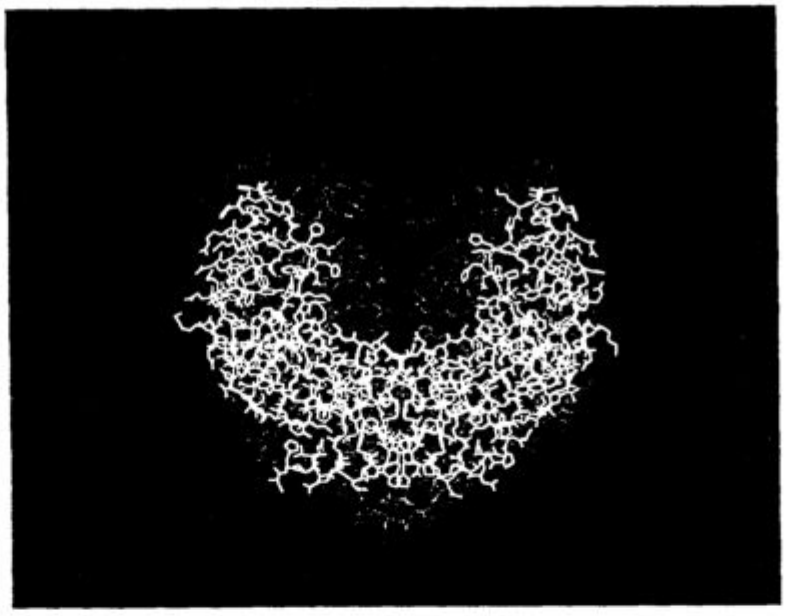
Οπως φαίνεται στον πίνακα 27, η στοίχιση του μονομερούς της R.*Pvu*II με αυτό της R.*Eco*RV, συγκριτικά με την στοίχιση των διμερών, παρά το γεγονός ότι επεκτείνεται σε μεγαλύτερο αριθμό καταλοίπων εμφανίζει μικρότερη rms

Πίνακας 28. Τρισδιάστατη στοίχιση των διμερών των R.PvuII και R.EcoRV (rms απόσταση 2.550Å). *Παρουσιάζονται τα καταλοιπα των οποίων τα C^{α} άτομα έρχονται σε γειτονικές στο χώρο θέσεις ύστερα από την στοίχιση καθώς και τα δευτεροταγή στοιχεία, της δομής, στα οποία ανήκουν. Από αυτά είναι υπογραμμισμένα τα όμοια κατάλοιπα.*

Μόριο	ΑΑ	Μονομερές Α	Μονομερές Β	Δευτερο ταχής δομή
PvuII EcoRV	1	35 NGGKLL 40 41 STIFEL 46	35 NGGKLLQ 41 41 STIFELF 47	αB αB
	2	71 SINID 75 93 TTYTN 97	71 SINID 75 93 TTYTN 97	βc βe
	3	78 KGFS 81 103 IKFT 106	78 KGFS 81 102 KIKF 105	βd βf
	4	102 AIYR 105 136 YVYT 139	102 AIYR 105 136 YVYT 139	βe βg
	5	143 KIPVT 147 191 SIHAH 195	141 NPKIPV 146 191 SIHAHY 196	αE αC

Πίνακας 29. Τρισδιάστατη στοίχιση των μονομερών των R.PvuII και R.EcoRV (rms απόσταση 2.083Å). *Παρουσιάζονται τα καταλοιπα των οποίων τα C^{α} άτομα έρχονται σε γειτονικές στο χώρο θέσεις ύστερα από την στοίχιση καθώς και τα δευτεροταγή στοιχεία, της δομής, στα οποία ανήκουν. Από αυτά είναι υπογραμμισμένα τα όμοια κατάλοιπα.*

Μόριο	ΑΑ	Στοίχιση Μονομερών Α	Στοίχιση Μονομερών Β	Δευτερο-ταχής δομή
R.*PvuII* R.*EcoRV*	1	38 KLLQ 41 41 STIF 44	38 KLLQ 41 41 STIF 44	αB αB
	2	56 GNDA 59 72 YPDF 75	56 GNDA 59 72 YPDF 75	βb βd
	3	67 YELKSINID 75 89 IDIKTTYTN 97	67 YELKSINID 75 89 IDIKTTYTN 97	βc βe
	4	78 KGESTHHH 85 103 IKETLGGY 110	78 KGESTHHH 85 103 IKETLGGY 110	βd βf
	5	96 QVPWIFAIYRG 106 130 AHWIIGYVYTR 140	96 QVPWIFAIYRG 106 130 AHWIIGYVYTR 140	βe βg
	6	110 EAIYRL 115 166 VKVFLQ 171	110 EAIYRL 115 166 VKVFLQ 171	βf βh
	7	124 YDKWERK 130 209 EDEFLDY 215	124 YDKWERKW 131 209 EDEFLDYW 216	αD αD
	8	140 NPKIP 145 188 IGSIH 193	141 NNPKIP 145 189 NIGSIH 193	βg βj
	9	148 YVMEH 152 198 DFVEG 202	148 YVMEH 152 198 DFVEG 202	αE αC

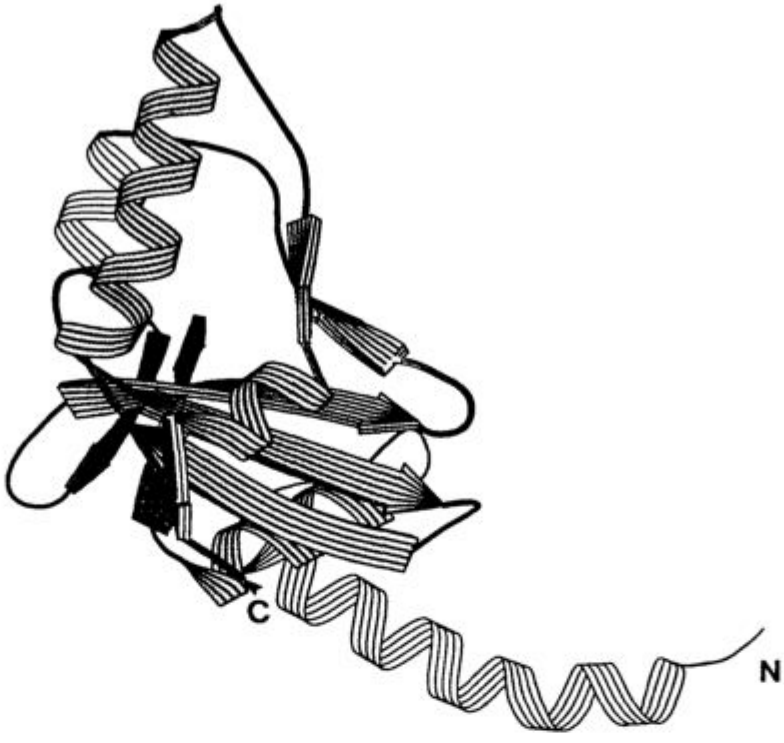


Εικόνα 47 . Ομοιότητα του συνολικού διπλώματος των δομών των R.PvuII και R.EcoRV. *Υπέρθεση του σκελετικού μοντέλου, της R.PvuII με κίτρινο χρώμα, με την ελεύθερη μορφή της EcoRV ενδονουκλεάσης με κόκκινο. Στο κέντρο η περιοχή πρόσδεσης του DNA της R.EcoRV.*

απόσταση των C^{α} ατόμων που στοιχίζονται. Το γεγονός αυτό μπορεί εύκολα να ερμηνευτεί με βάση την παρατήρηση ότι παρά την ομοιότητα των μονομερών των δύο μορίων ο διμερισμός διαφοροποιεί την σχετική τους θέση. Στην περίπτωση της R.*Eco*RV το άνοιγμα πρόσδεσης του DNA είναι σαφώς πιο κλειστό από ότι στην περίπτωση της R.*Pvu*II με αποτέλεσμα η στοίχιση του διμερούς να είναι ικανοποιητική μόνο στην περιοχή κοντά στον διπλό άξονα συμμετρίας των δύο μορίων.

Στους πίνακες 28-29 παρουσιάζονται τα κατάλοιπα που στοιχίζονται με rms απόσταση των C^{α} ατομών τους κάτω από 2.55Å (στοίχιση διμερούς) και 2.0Å (στοίχιση μονομερούς). Οπως μπορεί να παρατηρηθεί ελάχιστη είναι η ομοιότητα με βάση την αμινοξική αλληλουχία (υπογραμμισμένα κατάλοιπα, =14% ταυτότητα επί των στοιχισμένων αμινοξέων για την στοίχιση μονομερών) των καταλοίπων που στοιχίζονται δομικά.

Στην εικόνα 47 παρουσιάζονται οι δύο δομές ύστερα από χωρική υπέρθεση των δομικά ομόλογων περιοχών σε επίπεδο διμερούς της ελεύθερης πρωτεΐνης. Εύκολα διακρίνει κανείς αφενός την γενική ομοιότητα των δύο μορίων αλλά και το μεγαλύτερο κλείσιμο του χαρακτηριστικού "U" που εμφανίζει η R.*Eco*RV. Από την ίδια διαδικασία υπέρθεσης προέρχεται και η εικόνα 51 με την



Εικόνα 48. Χαρτογράφηση των δομικά ομόλογων περιοχών των R.PvuII και R.EcoRV πάνω στο μοντέλο της R.PvuII. *Οι σκιασμένες περιοχές έχουν τοπολογικά ισοδύναμες περιοχές στην δομή R.EcoRV. Από αυτές με ελαφριά σκίαση οι περιοχές που μετέχουν στην στοίχιση δίνοντας μία rms απόσταση των C^{α} ατόμων μικρότερη από 2.0Å. Με έντονη σκίαση περιοχές οι οποίες δεν υπερτίθενται με τις ισοδύναμες τους στην R.EcoRV.*

διαφορά ότι το μοντέλο παρουσιάζεται σε μορφή κορδέλας και μας επιτρέπει να διακρίνουμε τα δευτεροταγή στοιχεία της δομής των δύο πρωτεϊνών. Στην εικόνα αυτή γίνεται φανερό ότι η ομοιότητα των δύο μονομερών δεν περιλαμβάνει τις περιοχές διμερισμού των μορίων, είναι έντονη στην περιοχή της κεντρικής β-πτυχωτής επιφάνειας και επεκτείνεται στην περιοχή στο εσωτερικό των βραχιώνων του "U" που στην R.*Eco*RV έχει χαρακτηριστεί ως περιοχή αναγνώρισης του DNA.

Με βάση την τοπολογική ομοιότητα των δύο πρωτεϊνών προχωρήσαμε στην χαρτογράφηση του συνόλου των στοιχείων της δευτεροταγούς δομής για τα οποία παρατηρήθηκε αντιστοίχιση ανάμεσα στις δύο δομές. Το αποτέλεσμα που προέκυψε παρουσιάζεται στις εικόνες 48 και 50. Πάνω στο μοντέλο της R.*Pvu*II διακρίνονται τα ομόλογα στοιχεία ως σκιασμένες περιοχές. Η

Εικόνα 49. Υπέρθεση των μονομερών των R.EcoRV (ελεύθερη πρωτεΐνη) με κόκκινο - R.EcoRV (σύμπλοκο με συγγενή αλληλουχία) με μπλέ και R.PvuII με κίτρινο. *Διακρίνονται οι διαφοροποιήσεις μεταξύ των δύο μορφών της R.EcoRV στις περιοχές των βρόγχων όπως και στην περιοχή διμερισμού του μορίου. Επίσης είναι εμφανείς οι τρείς βρόγχοι της R.EcoRV που μετέχουν στην πρόσδεση του DNA.*

αντιστοίχιση των στοιχείων της δευτεροταγους δομής δεν είναι συνεχόμενη πάνω στην αλληλουχία των δύο πρωτεϊνών (εικόνα 50) υποδηλώνοντας ότι, αν η ομοιότητα είναι αποτέλεσμα αποκλείνουσας εξέλιξης, τότε έχουν συμβεί γεγονότα επαναδιάταξης των δευτεροταγών στοιχείων τους.

Η επίλυση της δομής της R.*Eco*RV σε μορφή συμπλόκου με την συγγενή αλληλουχία DNA επέτρεψε την ταυτοποίηση των περιοχών της που είναι υπεύθυνες για την αναγνώριση της αλληλουχίας και για την κατάλυση της πέψης του DNA. Το γεγονός της τοπολογικής ομοιότητας των δύο ενζύμων που περιλαμβάνει τις ταυτοποιημένες περιοχές της R.*Eco*RV μας επέτρεψε την άμεση σύγκριση των περιοχών αυτών.

Δύο βρόγχοι της R.*Eco*RV χαρακτηρισμένοι ως R (recognition) (182-187) και Q (Gln68-His71) {Winkler, 1993} σχηματίζουν το κύριο μέρος των αλληλεπιδράσεων με το συγγενές DNA. Από αυτούς ο R βρόγχος παρουσιάζεται υπεύθυνος για τις ειδικές αλληλεπιδράσεις ενώ ο δεύτερος αλληλεπιδρά κύριως με την ραχοκοκκαλιά του DNA ενώ σχηματίζει την μοναδική πιθανή ειδική αλληλεπίδραση, με βάση, στο μικρό αυλάκι του DNA. Η αντίστοιχη περιοχή του Q

Εικόνα 49. Υπέρθεση των μονομερών των R.EcoRV (ελεύθερη πρωτεΐνη) με κόκκινο - R.EcoRV (σύμπλοκο με συγγενή αλληλουχία) με μπλέ και R.PvuII με κίτρινο. *Διακρίνονται οι διαφοροποιήσεις μεταξύ των δύο μορφών της R.EcoRV στις περιοχές των βρόγχων όπως και στην περιοχή διμερισμού του μορίου. Επίσης είναι εμφανείς οι τρείς βρόγχοι της R.EcoRV που μετέχουν στην πρόσδεση του DNA.*

αντιστοίχιση των στοιχείων της δευτεροταγους δομής δεν είναι συνεχόμενη πάνω στην αλληλουχία των δύο πρωτεϊνών (εικόνα 50) υποδηλώνοντας ότι, αν η ομοιότητα είναι αποτέλεσμα αποκλείνουσας εξέλιξης, τότε έχουν συμβεί γεγονότα επαναδιάταξης των δευτεροταγών στοιχείων τους.

Η επίλυση της δομής της R.*Eco*RV σε μορφή συμπλόκου με την συγγενή αλληλουχία DNA επέτρεψε την ταυτοποίηση των περιοχών της που είναι υπεύθυνες για την αναγνώριση της αλληλουχίας και για την κατάλυση της πέψης του DNA. Το γεγονός της τοπολογικής ομοιότητας των δύο ενζύμων που περιλαμβάνει τις ταυτοποιημένες περιοχές της R.*Eco*RV μας επέτρεψε την άμεση σύγκριση των περιοχών αυτών.

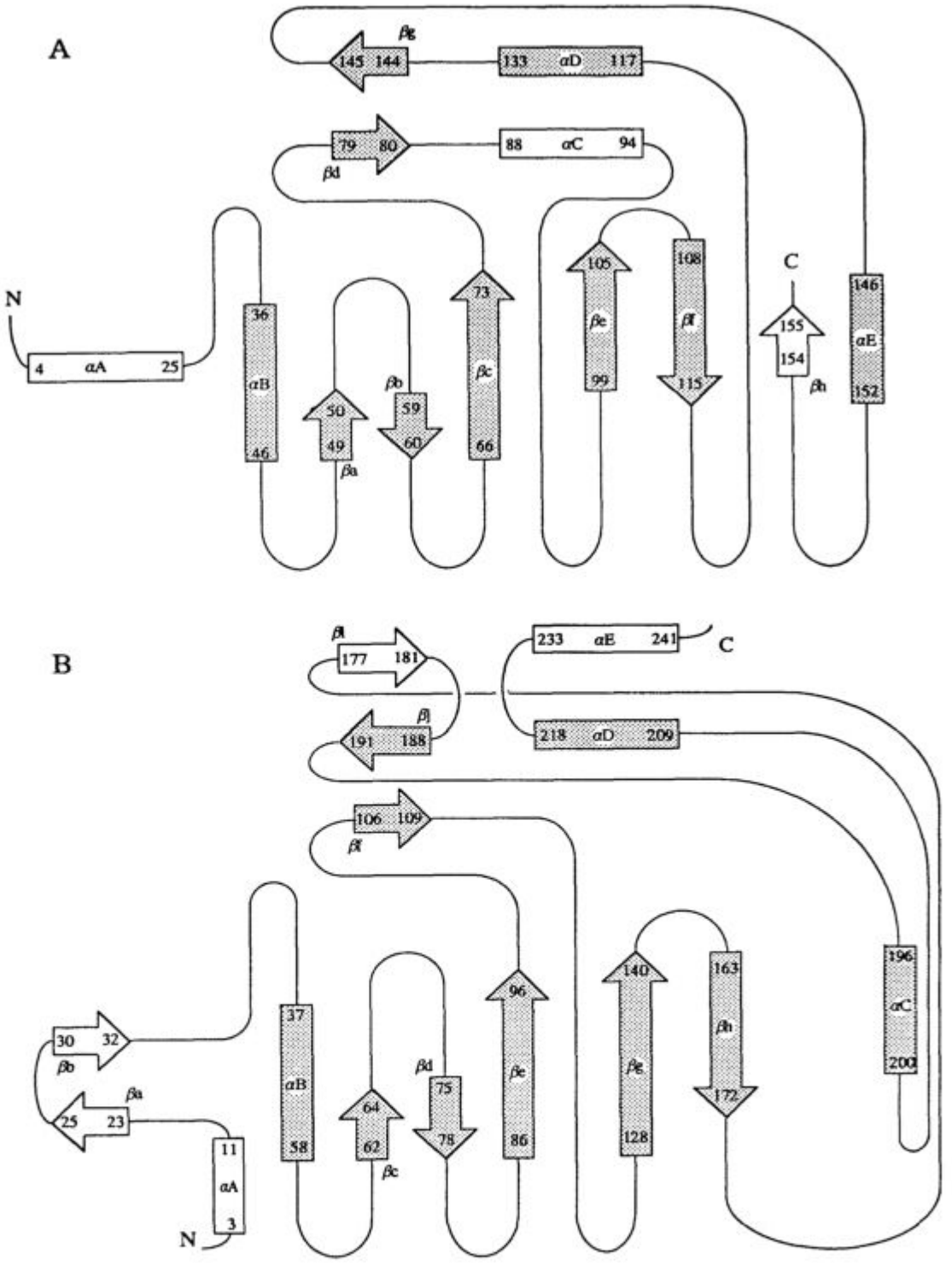
Δύο βρόγχοι της R.*Eco*RV χαρακτηρισμένοι ως R (recognition) (182-187) και Q (Gln68-His71) {Winkler, 1993} σχηματίζουν το κύριο μέρος των αλληλεπιδράσεων με το συγγενές DNA. Από αυτούς ο R βρόγχος παρουσιάζεται υπεύθυνος για τις ειδικές αλληλεπιδράσεις ενώ ο δεύτερος αλληλεπιδρά κύριως με την ραχοκοκκαλιά του DNA ενώ σχηματίζει την μοναδική πιθανή ειδική αλληλεπίδραση, με βάση, στο μικρό αυλάκι του DNA. Η αντίστοιχη περιοχή του Q

βρόγχου στο μόριο της R.*Pvu*II συνδέει τους βa και βb β-κλώνους κατά ανάλογο τρόπο με αυτόν της R.*Eco*RV. Η περιοχή R στην R.*Eco*RV συνδέει τους β-κλώνους βi και βj, από αυτούς μόνο ο δεύτερος έχει ισοδύναμο του στην δομή της R.*Pvu*II ενώ ο β-κλώνος βi έχει αντικατασταθεί από την α-έλικα αC. Το γεγονός της διαφοροποίησης της R περιοχής, παρά την δομική συντήρηση λίγο ως πολύ των γειτονικών δευτεροταγών στοιχείων, αφενός υποδηλώνει την διαφοροποίηση που απαιτείται για την αναγνώριση διαφορετικών αλληλουχιών και αφεταίρου, σε συνδυασμό με το γεγονός ότι ο Q βρόγχος δεν έχει ορισμένη δομή στις ελεύθερες R.*Eco*RV και R.*Pvu*II, την ευκινησία που απαιτείται κατά την διαδικασία της πρόσδεσης.

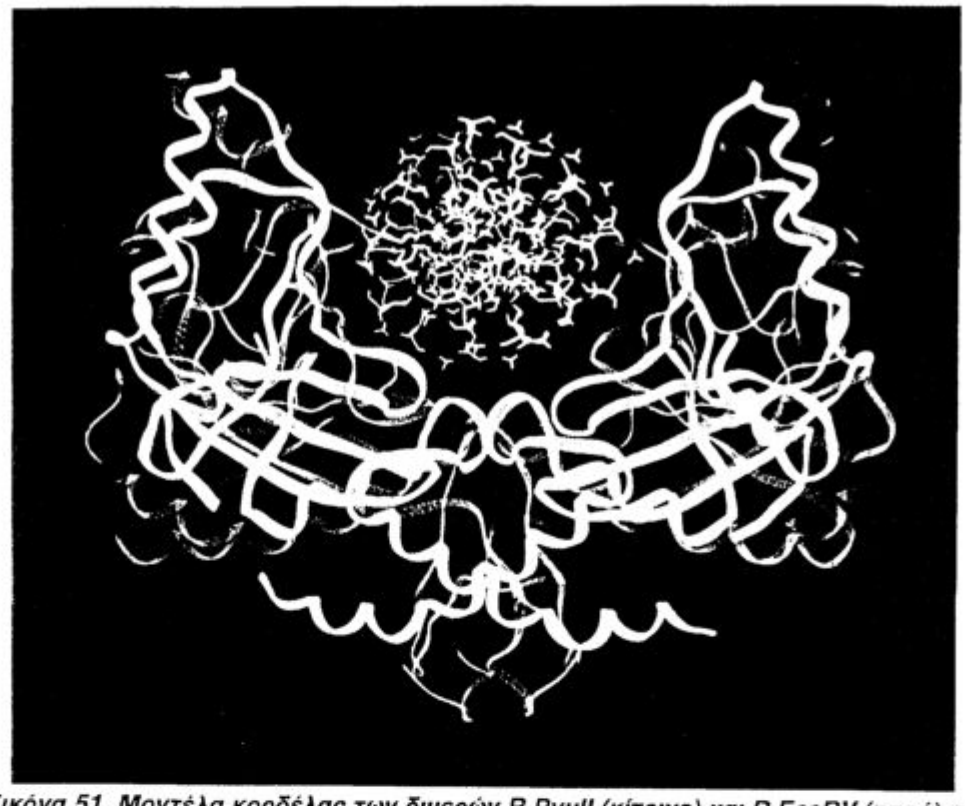
Μια δεύτερη περιοχή της R.*Eco*RV η οποία αλληλεπιδρά εκτεταμένα με την ραχοκοκκαλιά του DNA περιλαμβάνει την καταλυτική περιοχή και ένα τμήμα της περιοχής διμερισμού. Στην περιοχή αυτή σχηματίζονται 10 υδρογονικοί δεσμοί μεταξύ πρωτεΐνης και της ραχοκοκκαλιάς του DNA, 6 από αυτούς σχηματίζονται από τα κατάλοιπα 92-95 που ανήκουν στον β-κλώνο βe, μια περιοχή που συμμετέχει στον πυρήνα των δομικά ομόλογων περιοχών των δύο μορίων (πίνακας 29, περιοχή 3). Το πρώτο κατάλοιπο της περιοχής αυτής, Lys92, είναι ένα από τα ελάχιστα συντηρημένα κατάλοιπα στις αλληλουχίες των δομικά ομόλογων περιοχών.

Στο καταλυτικό κέντρο της R.*Eco*RV οι πλευρικές αλυσίδες των 4 κατάλοιπων Asp74, Asp90, Lys92 και Asn188 έρχονται σε κοντινή απόσταση από την ενεργοποιημένη φωσφορική ομάδα που θα αποτελέσει την τομή της αλυσίδας του DNA. Και τα τέσσερα αυτά κατάλοιπα διατηρούν τη θέση τους στην δομή του μονομερούς της R.*Pvu*II δείχνοντας απόλυτη συντήρηση του ενεργού κέντρου των δύο περιοριστικών τόσο δομικά όσο και από πλευράς σύνθεσης των πλευρικών αλυσίδων που μετέχουν στο σχηματισμό του, αφού τα αντίστοιχα αμινοξέα στην περίπτωση της R.*Pvu*II είναι τα Asp58, Glu68, Lys70 και Asn141 (πίνακας 29). Τα δύο καρβοξύλια του καταλυτικού κέντρου μαζί με μία γειτονική φωσφορική ομάδα του DNA έχει υποτεθεί ότι σχηματίζουν την περιοχή πρόσδεσης του Mg^{2+}. Η αμινομάδα της Lys92 δημιουργεί γέφυρα ανάμεσα στα φωσφορικές ομάδες 0, +1 του DNA και δημιουργεί ένα τρίτο υδρογονικό δεσμό με το C=O της ραχοκοκκαλιάς της Thr106 της οποίας το αντίστοιχο αμινοξύ στην R.*Pvu*II είναι η Ser81 ή η Thr82.

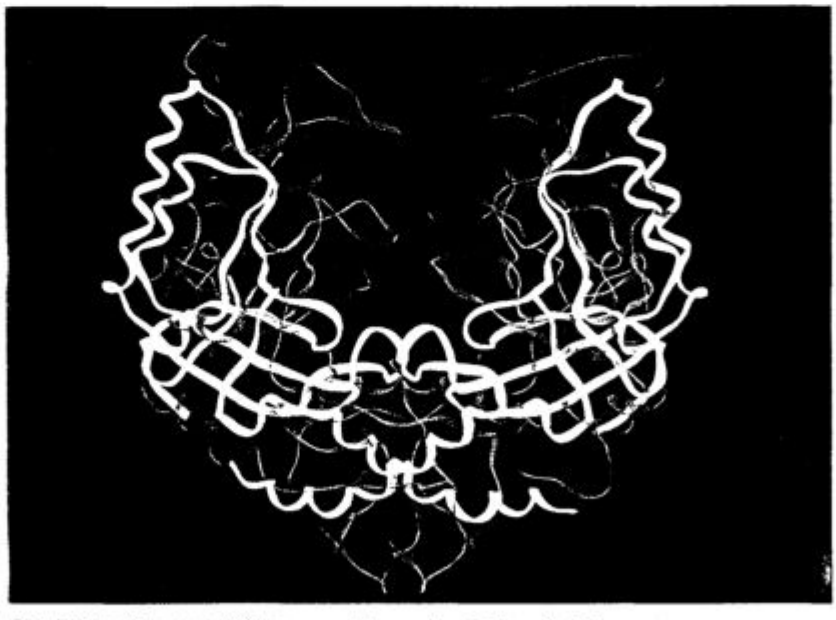
Ανάλογη ομοιότητα των ενεργών κέντρων έχει παρατηρηθεί και ανάμεσα στις R.*Eco*RI και R.*Eco*RV παρά την απουσία άλλων κοινών δομικών χαρακτηριστικών. Ετσι η δομή της *Pvu*II ενδονουκλεάσης έρχεται να επιβεβαιώσει την υπόθεση της ύπαρξης κοινού ενεργού κέντρου για τα τύπου II περιοριστικά ένζυμα παρά τις όποιες διαφορές τους στον τρόπο πρόσδεσης και αναγνώρισης του DNA.



Εικόνα 50. Τοπολογικό διάγραμμα των R.PvuII και R.EcoRV. *Σκιασμένες οι τοπολογικά ισοδύναμες περιοχές των δύο μορίων. PvuII(A) και R.EcoRV(B). Η αρίθμηση αφορά το πρώτο και τελευταίο κατάλοιπο των δευτεροταγών στοιχείων και N και C τα αμίνο- και καρβόξυ-τελικά άκρα των πρωτεινών. Ο ορισμός των δευτεροταγών στοιχείων της R.EcoRV έγινε με βάση τους Winkler et al. (Winkler et al. 1993].*



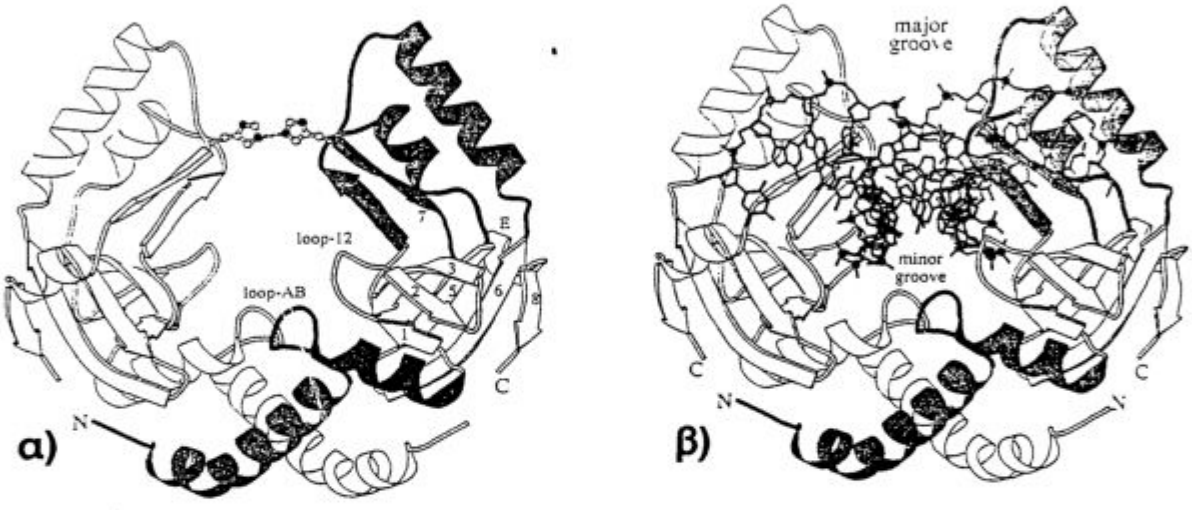
Εικόνα 51. Μοντέλα κορδέλας των διμερών R.PvuII (κίτρινο) και R.EcoRV (καφέ) σε σύμπλοκο με μή συγγενές DNA ύστερα από υπέρθεση. *Ο άξονας της διπλής έλικας του DNA είναι κάθετος στο επίπεδο της φωτογραφίας. Διακρίνονται οι βρόγχοι της R.EcoRV που προσεγκίζουν το μεγάλο αυλάκι του DNA.*



Εικόνα 52. Μοντέλα κορδέλας των διμερών R.PvuII (κίτρινο) και R.EcoRV στην ελεύθερη από DNA μορφή της (κόκκινο) ύστερα από υπέρθεση. *Αντίστοιχη εικόνα με την εικόνα 49 με διαφορετικού τύπου παράσταση ώστε να διακρίνονται τα δευτεροταγή στοιχεία των δομών.*

Β10.2. Σύγκριση της ελεύθερης R.*Pvu*II με την πρωτεΐνη σε σύμπλοκο με συγγενές DNA.

Η κρυστάλλωση {Balendiran et al., 1994} και ο προσδιορισμός της δομής του συμπλόκου της *Pvu*II ενδονουκλεάσης με την συγγενή αλληλουχία του DNA (εικόνα 53) {Cheng et. al., 1994} και η παραχώρηση των συντεταγμένων της δομής, μας επέτρεψε την σύγκριση των δύο δομών και την εξαγωγή συμπερασμάτων σχετικά με τις δομικές μεταβολές που παρατηρούνται, στην πρωτεΐνη, κατά την διαδικασία πρόσδεσης της στο DNA.



Εικόνα 53. Σχηματική παράσταση της R.PvuII στη μορφή του συμπλόκου με συγγενές DNA. {Cheng et al., 1994}. *α) Διακρίνονται οι His85 των δύο μονομερών που στην δομή της ελεύθερης πρωτεΐνης απέχουν 26Å. β) Η ίδια εικόνα του μορίου, παρουσία του DNA, σημειώνονται το μικρό (minor groove) και το μεγάλο αυλάκι (major groove) του DNA. Και στις δύο εικόνες N και C το αμινοτελικό και το καρβόξυτελικό άκρο της πρωτεΐνης αντίστοιχα*

Οι δομικές αλλαγές που επάγονται από την πρόσδεση στο DNA.

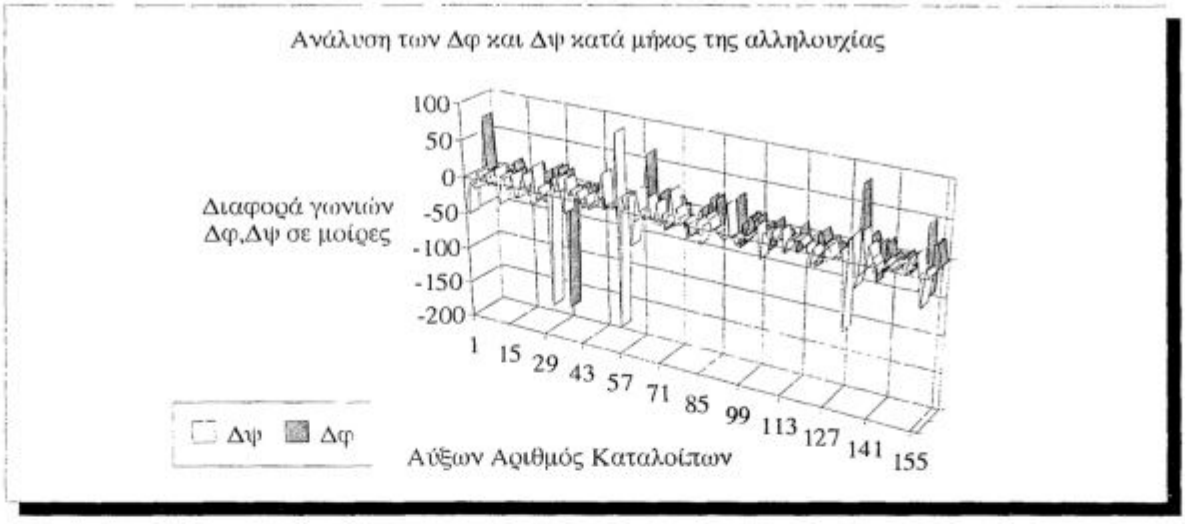
Μια πρώτη στοίχιση στο χώρο των δύο δομών βασισμένη κυρίως στην περιοχή διμερισμού (κατάλοιπα 4-45) προκείμενου οι δύο διαμορφώσεις της πρωτεΐνης να αποκτήσουν την ίδια σχετική διεύθυνση στο χώρο έδειξε την ύπαρξη σημαντικών δομικών διαφοροποιήσεων (εικόνες 55α, 55β). Στην δομή του συμπλόκου της πρωτεΐνης η His85 του κάθε μονομερούς βρίσκεται σε απόσταση σχηματισμού υδρογονικού δεσμού με την His85 του άλλου μονομερούς της

πρωτεΐνης (εικόνα 53α). Καθώς ο δεσμός αυτός βρίσκεται στην πλευρά στην οποία το μόριο του DNA εκτίθεται στον διαλύτη (μεγάλο αυλάκι) είναι ενδεικτικός του ιδιαίτερα κλειστού πακεταρίσματος πρωτεΐνης-DNA της R.*Pvu*II. Στην συνέχεια προσπαθήσαμε να εντοπίσουμε τα στοιχεία της δομής της πρωτεΐνης που υπόκεινται σε αλλαγές της διαμόρφωσης τους και είναι υπεύθυνα για τις παρατηρούμενες διαφορές στην συνολική διαμόρφωση. Η σύγκριση των δύο μορφών της πρωτεΐνης πραγματοποιήθηκε α) με βάση την σύγκριση των φ, ψ γωνιών της κύριας αλυσίδας της πρωτεΐνης και β) με τρισδιάστατη στοίχιση των δομών και σύγκριση των rms αποστάσεων των ισοδύναμων C^{α} ατόμων.

Οι φ και ψ γωνίες της δομής του συμπλόκου όπως και της ελεύθερης πρωτεΐνης υπολογίστηκαν από τα μοντέλα των δομών με το πρόγραμμα DSSP{Kabsch & Sander, 1983}. Στην συνέχεια επεξεργαστήκαμε τις τιμές των φ,ψ γωνιών με το πρόγραμμα EXCELL προκειμένου να βρούμε τις διαφορές τους για κάθε κατάλοιπο. Στην εικόνα 54 παρουσιάζονται διαγραμματικά οι διαφορές των φ και ψ γωνιών για όλα τα αμινοξέα των μονομερών A (όπως αυτά ορίζονται στις αντίστοιχες εγγραφές στην βάση δεδομένων Brookhaven) των δύο μορίων. Γίνεται φανερό ότι οι μεγάλες διαφορές των φ,ψ γωνιών ανάμεσα στις δύο δομές (πίνακας 30) εντοπίζονται σε ένα περιορισμένο αριθμό καταλοίπων ενώ οι πλειοψηφία των καταλοίπων εμφανίζει διαφορές μικρότερες των 40°. Οι μεγαλύτερες διαφοροποιήσεις των φ και ψ γωνιών εντοπίζονται σε τρία σημεία της αλληλουχίας της πρωτεΐνης : α) Στην περίπτωση των καταλοίπων Gln33-Asp34 η διαφοροποίηση (όπως φαίνεται ύστερα από οπτική επίβλεψη των μοντέλων) οφείλεται σε αντιδιαμετρική τοποθέτηση του C=O της κύριας αλυσίδας του Gln33 καταλοίπου και είναι παρούσα μόνο στην περίπτωση του μονομερούς A. β) Η περιοχή Gly53-Asp58 παρουσιάζεται αποδιοργανωμένη στους χάρτες

Πίνακας 30. Διαφοροποίηση των φ-ψ γωνιών ανάμεσα στην ελεύθερη R.*Pvu*II και στο συμπλόκο της με το DNA. Σημειώνονται οι διαφορές που είναι μεγαλύτερες από 40° και παρουσιάζονται και στα δύο μονομερή. Τα (A) και (B) χαρακτηρίζουν την σύγκριση μεταξύ των μονομερών A και B αντίστοιχα των μοντέλων.

Κατάλοιπα	Δφ	Δψ
Ser2		-53(A), -140(B)
Gly53	-128(A)	-152(A), 65(B)
Asn57		99(A), -126(B)
Asp58		-164(A),-151(B)
Asn62		-46(A),-42(B)
Ser133	61(B)	-94(A)
Asp134	84(A)	44(B)



Εικόνα 54. Ανάλυση των Δφ και Δψ κατά μήκος της αλληλουχίας.

ηλεκτρονιακής πυκνότητας της ελεύθερης πρωτεΐνης και σε μικρότερο βαθμό σε αυτήν του συμπλόκου με το DNA. Η διαμόρφωση της είναι σημαντικά διαφορετική στις δύο δομές ενώ στην περίπτωση της ελεύθερης πρωτεΐνης περιλαμβάνει τα δύο αμινοξέα Arg54 και Glu55 των οποίων η ηλεκτρονιακή πυκνότητα απουσίαζε από τους χάρτες και γ) Τέλος η περιοχή Ser133-Asp134 ανήκει στον δεύτερο βρόγχο της πρωτεΐνης με ψηλούς παράγοντες θερμοκρασίας και είναι σχετικά αποδιοργανομένη στους χάρτες ηλεκτρονιακής πυκνότητας. Σε όλες τις περιπτώσεις οι σημαντικές διαφοροποιήσεις των φ, ψ γωνιών αντιπροσωπεύουν τοπικές διαφοροποιήσεις ανάμεσα στούς βρόγχους των δύο δομών και δεν ερμηνεύουν τις σημαντικές αλλαγές διαμόρφωσης που παρατηρούνται στο συνολικό δίπλωμα του μορίου (εικόνα 55α, β - εικόνα56α, β).

Στον πίνακα 31 παρουσιάζονται τα αποτελέσματα της τρισδιάστατης στοίχισης μεταξύ της ελεύθερης πρωτεΐνης και του συμπλόκου με το DNA. Σύμφωνα με τα αποτελέσματα αυτά το μονομερές αποτελείται από δύο περιοχές: Pro4-Leu51 και Asp58-Tyr157 οπου κάθε μία χωριστά μπορεί να έρθει σε ταύτιση με την αντίστοιχη του συμπλόκου (0.897Å και 0.678Å rms απόσταση αντίστοιχα). Η περιοχή των καταλοίπων Pro52-Asn57 περιλαμβάνει τα αμινοξέα Arg54 και Glu55 που απουσιάζουν από το μοντέλο της ελεύθερης πρωτεΐνης και όπως αναφέρθηκε αποτελεί μία περιοχή με χαμηλό βαθμό ταξης. Ακόμη και εξαιρώντας αυτήν την περιοχή, προσπάθεια για τρισδιάστατη στοίχιση ολόκληρου του μονομερούς, αποτυγχάνει, δίνοντας 2.961Å rms απόσταση των C^{α} ατόμων, υποδηλώνοντας διαφοροποίηση της σχετικής θέσης της περιοχής πρόσδεσης του DNA σε σχέση με την περιοχή διμερισμού της πρωτεΐνης.

Στην τέταρτη σειρά του πίνακα παρουσιάζεται η βελτιστοποίηση της στοίχισης του συνόλου των καταλοίπων του μονομερούς. Γίνεται φανερό ότι παρά το γεγονός α) ότι η περιοχή Pro4-Leu51 διατηρεί την δομή της στα δύο μοντέλα και β) και ότι η σχετική θέση της υποπεριοχής αυτής σε σχέση με το υπόλοιπο μόριο μεταβάλλεται, υπάρχουν τμήματα της περιοχής αυτής (κατάλοιπα Tyr19-Asn29 και Asn35-Val50) που συνεχίζουν να διατηρούν την θέση τους σε σχέση με την C-τελική περιοχή Asp58-Tyr157. Αυτό θα μπορούσε να συμβαίνει στην περίπτωση που η σχετική μετακίνηση των δύο υποπεριοχών είναι αποτέλεσμα περιστροφής γύρω από άξονα που ορίζεται από τις περιοχές αυτές. Κάτι τέτοιο θα ήταν αποτέλεσμα διαφοροποίησης των φ και ψ γωνιών των καταλοίπων στα όρια των περιοχών που διατηρούν την θέση τους σε σχέση με την C-τελική περιοχή. Κάτι τέτοιο δεν είναι εμφανές σε πρώτη προσέγγιση, μιάς και όπως αναφέρθηκε νωρίτερα οι μεγάλες διαφορές που παρατηρούνται στα κατάλοιπα Gln33 και Asp34 περιορίζονται στο ένα μονομερές και είναι αποτέλεσμα της αντιστροφής του C=O του Gln33 κατά την ερμηνεία του χάρτη ηλεκτρονιακής πυκνότητας και δεν έχει αποτέλεσμα στην διαμόρφωση της ευρύτερης περιοχής. Όμως, προσεκτική μελέτη των δύο μοντέλων ύστερα από στοίχιση των C-τελικών περιοχών των μονομερών, έδειξε ότι η διαφοροποίηση των μοντέλων είχε πράγματι ως σημείο εκκίνησης την περιοχή των αμινοξέων Gly36,Gly37 και Lys38 για τα οποία, αν και δεν εμφανίζεται μεγάλη διαφοροποίηση στις φ,ψ γωνίες τους ανάμεσα στα δύο μοντέλα, υποθέσαμε ότι λειτουργώντας συνεργατικά έχουν το παρατηρούμενο αποτέλεσμα. Για να διαπιστώσουμε την ορθότητα της υπόθεσης αυτής πραγματοποιήσαμε το θεωρητικό πείραμα της "μεταλαγής" της δομής του συμπλόκου (μοντέλα Gly36 και GGK) αποδίδοντας στα αμινοξέα αυτά, στο μοντέλο του συμπλόκου, τις τιμές των φ και ψ γωνιών που έχουν στην ελεύθερη πρωτεΐνη. Το αποτέλεσμα ήταν σημαντικές αλλαγές στη διαμόρφωση του μοντέλου φέρνοντας το πολύ κοντά στην διαμόρφωση της ελεύθερης πρωτεΐνης. Η rms απόσταση για τα C^{α} άτομα του μονομερούς A, και για το Gly36 μοντέλο, από 2.961Å μεταξύ ελεύθερης πρωτεΐνης και συμπλόκου γίνεται 2.186Å, και 1.474Å για το μοντέλο με τις αλλαγές των φ και ψ και των τριών καταλοίπων Gly36, Gly37, Lys38 (GGK) (πίνακας 31). Τα κατάλοιπα Gly36-Lys38 ανήκουν στην πρώτη στροφή της έλικας αB και ο άξονας σχηματισμού του διμερούς διέρχεται σε κοντινή απόσταση από αυτά. Η θέση τους στην κέντρο της περιοχής (σχισμής) πρόσδεσης του DNA επιτρέπει την αλληλεπίδραση τους (και των γειτονικών τους αμινοξέων) με τις κεντρικές βάσεις της όποιας αλληλουχίας του DNA που βρίσκεται προσδεδεμένο ειδικά η όχι. Η περιοχή αυτή της πρωτεΐνης είναι η μόνη που, στην ελεύθερη μορφή του ενζύμου, μπορεί να αλληλεπιδράσει συμμετρικά με το μόριο του DNA. Τίθεται λοιπόν το ερώτημα κατά πόσον η περιοχή των καταλοίπων Gly36-Lys38 και η γειτονική της,

Πίνακας 31. Σύγκριση της δομής του συμπλόκου και των μοντέλων GGK, Gly36 με αυτήν της ελεύθερης πρωτεΐνης βασισμένη στις rms αποστάσεις C^{α} ατόμων περιοχών καταλοίπων των δύο μοντέλων. Τα δύο μονομερή των μοντέλων χαρακτηρίζονται ως A και B. Οπου δεν υπάρχει σχετική αναφορά η σύγκριση έχει πραγματοποιηθεί για κάθε μονομερές ξεχωριστά.

Περιοχή Κατάλοιπων	rms απόστασεις για το μονομερές A (Å)	rms απόστασεις για το μονομερές B (Å)
4-51	0.897	1.052
58-157	0.678	0.618
2-53, 56-157	2.961	2.682
19-29, 35-50, 58-157	1.196	1.092
A2-A53, A56-A157, B2-B53, B53-B157	6.180	-
Gly36 2-53, 56-157	2.186	2.599
Gly36 A2-B157	3.264	-
GGK 2-53, 56-157	1.852	1.430
GGK A2-A53, A56-A157, B2-B53, B53-B157	2.222	-
GGK 4-51, 58-157	1.474	1.045 (περιοχή 2-50,58-157)
GGK A5-A51, A58-A157 B12-B51, B58-B157	1.809	-

Πίνακας 31α. Σύγκριση της δομής των μονομερών των μοντέλων της ελεύθερης R.*Pvu*II και του συμπλόκου με την συγγενή αλληλουχία. Οι τιμές των rms αποστάσεων μεταξύ μονομερών του ίδιου μορίου παρατείθενται για λόγους εκτίμησης των αποτελεσμάτων του πίνακα 31.

Περιοχή Κατάλοιπων	rms απόστασεις (Å)
2-53, 56-157 συγκριση μεταξύ των μονομερών της ελεύθερης R.*Pvu*II	0.789
2-53, 56-157 συγκριση μεταξύ των μονομερών του μοντέλου του συμπλόκου.	0.548

του βρόγχου που συνδέει τις α-έλικες αA και αB (κατάλοιπα 26-35), λειτουργούν ως αισθητήρες της παρουσίας του DNA και επάγουν την αλλαγή της συνολικής διαμόρφωσης του μορίου.

Οπως αναφέραμε στο κεφάλαιο της σύγκρισης της δομής της R.*Pvu*II με αυτήν της R.*Eco*RV το κατάλοιπο Asp58 φαίνεται να είναι δομικά ομόλογο με ένα

από τα αμινοξέα που μετέχουν στην δημιουργία του καταλυτικού κέντρου του μορίου. Στην δομή του συμπλόκου της R.*Pvu*II μια δεύτερη όξινη αλυσίδα αυτή του Glu55 πλησιάζει την περιοχή πρόσδεσης του Mg^{2+}. Τα δύο αμινοξέα είναι εξίσου υποψήφια για την συμμετοχή τους στο καταλυτικό κέντρο. Το Glu55 κατάλοιπο στην δομή της ελεύθερης μορφής του ενζύμου βρίσκεται σε αποδιοργανωμένη μορφή και δεν είναι ορατό στην ηλεκτρονιακή πυκνότητα, αλλά η προβλεπόμενη θέση του με βάση την συνέχεια της αλυσίδας της ραχοκοκκαλιάς της πρωτεΐνης το τοποθετεί σε μεγάλη απόσταση από την θέση που κατέχει στην δομή του συμπλόκου σχετικά με τα υπόλοιπα κατάλοιπα του ενεργού κέντρου. Ολόκληρη η περιοχή που περιλαμβάνει το Glu55 κατάλοιπο από την Pro51 εως το Asn57 αποτελεί την μόνη περιοχή της πρωτεΐνης που αποκτά διαφορετική τοπική διαμόρφωση κατα το πέρασμα από την ελεύθερη πρωτεΐνη στο σύμπλοκο με το DNA. Τέλος η ίδια περιοχή εμφανίζει ψηλούς B-παράγοντες και στις δύο δομές με μεγαλύτερους σε αυτήν της ελεύθερης πρωτεΐνης. Στην περίπτωση όπου το Glu55 κατάλοιπο αποτελεί πράγματι τμήμα του καταλυτικού κέντρου τα προηγούμενα συνηγορούν στην υπόθεση ότι το ενεργό κέντρο της πρωτεΐνης δομικά σχηματίζεται μόνο μετά την πρόσδεση στο μόριο του DNA και πιθανά κατά την πρόσδεση στην ειδική αλληλουχία αναγνώρισης.

Συνοψίζοντας η πρόσδεση της πρωτεΐνης στο DNA στόχο, συμβαδίζει με δύο σημαντικές διαφοροποιήσεις στην δομή της :

α) ολόκληρη η περιοχή πρόσδεσης του DNA σε κάθε μονομερές πραγματοποιεί μία περιστροφή με άξονα που διέρχεται από τα κατάλοιπα Gly36,Gly37 και Lys38 έχοντας ως αποτέλεσμα το κλείσιμο των βραχιόνων και την προσέγγιση, των καταλοίπων που μετέχουν στην αναγνώριση, στον άξονα του DNA.

β) η περιοχή των αμινοξέων 52-57 αποκτά μερική τάξη και σε συνδυασμό με την περιστροφή της υποπεριοχής πρόσδεσης του DNA στην οποία ανήκει, πλησιάζει τον ενεργοποιημένο φωσφοδιεστερικό δεσμό.

ΔΙΑΓΩΝΙΣΜΑ 4ο

ΖΗΤΗΜΑ 1ο

α) Αν η συνάρτηση $f:\mathbb{R}\to\mathbb{R}$ είναι συνεχής στο $x_0=1$ και $\lim\limits_{x\to 1}\frac{f(x)-x^3}{x^2-1}=7$ να δειχθεί ότι η f είναι παραγωγίσιμη στο $x_0=1$

β) Έστω $f:\mathbb{R}\to\mathbb{R}$ μια παραγωγίσιμη συνάρτηση τέτοια ώστε για κάθε $x\in\mathbb{R}$ να ισχύει

$$f(x^3+x)=x^3$$

Να βρείτε την εξίσωση της εφαπτομένης της γραφικής παράστασης της f στο σημείο της $M(2,f(2))$.

ΖΗΤΗΜΑ 2ο

α) Θεωρούμε κύκλο (O,R) με διάμετρο AB. Κινητό σημείο M διαγράφει το ημικύκλιο $\overset{\frown}{AB}$. Να βρείτε τον ρυθμό μεταβολής της απόστασης MK του σημείου M από την διάμετρο AB ως προς την γωνία $\theta=A\hat{B}M$, όταν $MK=R$.

β) Αν C_1, C_2 είναι γραφικές παραστάσεις των συναρτήσεων

$$f(x)=x+1-\frac{1}{x} \qquad \text{και} \qquad g(x)=2x^2+6x+\alpha$$

αντιστοίχως, να βρείτε την τιμή του $\alpha\in\mathbb{R}$ ώστε η εφαπτόμενη της C_1 στο σημείο $M(1,f(1))$ να εφάπτεται και στη C_2.

ΖΗΤΗΜΑ 3ο

α) Έστω $f(x)$ ένα πολυώνυμο βαθμού $\nu\geq 2$. Να δειχθεί ότι το $f(x)$ έχει παράγοντα το $(x-\rho)^2$ αν και μόνο αν ισχύει $f(\rho)=f'(\rho)=0$ $(\rho\in\mathbb{R})$.

β) 1. Να δειχθεί ότι το $(x-1)^2$ είναι παράγοντας του πολυωνύμου

$$f(x)=\nu\cdot x^{\nu+1}-x^{\nu-1}-(\nu^2+1)x+\nu^2-\nu+2,\ \nu\in\mathbb{N} \text{ με } \nu\geq 2.$$

2. Να βρείτε τις τιμές των α,β για τις οποίες το $(x-1)^2$ είναι παράγοντας του $\varphi(x)=\alpha x^{2\nu}+\beta x^{\nu-1}+4$, $\nu\in\mathbb{N}$ με $\nu\geq 2$

ΖΗΤΗΜΑ 4ο

α) Έστω $f(x)$ ένα πολυώνυμο 3ου βαθμού με πραγματικούς συντελεστές και $\rho_1,\rho_2,\rho_3\in\mathbb{R}_+^*$ οι ρίζες του. Αν $x_0=\rho_1+\rho_2+\rho_3$ να δειχθεί ότι:

1. $\dfrac{f'(x)}{f(x)}=\dfrac{1}{x-\rho_1}+\dfrac{1}{x-\rho_2}+\dfrac{1}{x-\rho_3}$, $x\in\mathbb{R}\setminus\{\rho_1,\rho_2,\rho_3\}$

2. $x_0\cdot\dfrac{f'(x_0)}{f(x_0)}\geq\dfrac{9}{2}$

β) Έστω συνάρτηση $f:\mathbb{R}\to\mathbb{R}$ με

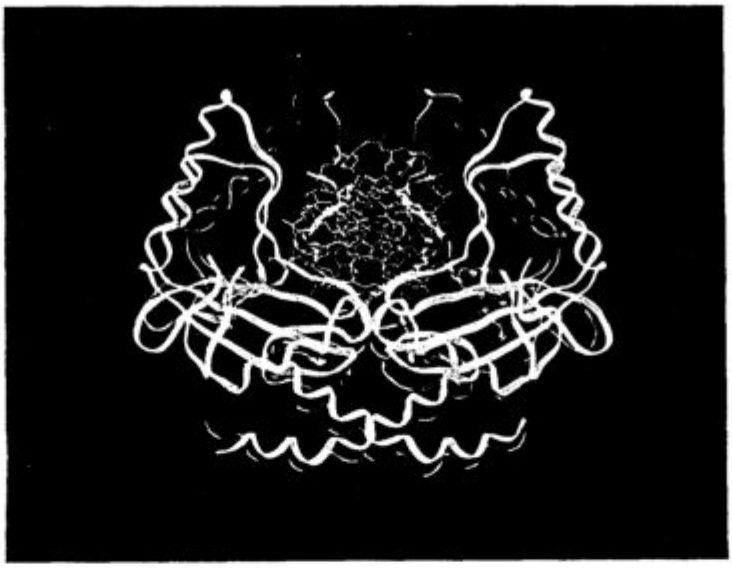
$$f(x+y)=f(x)\cdot\sigma\upsilon\nu y+f(y)\cdot\sigma\upsilon\nu x \text{ για κάθε } x,y\in\mathbb{R}.$$

1. Να δειχθεί ότι $f(0)=0$

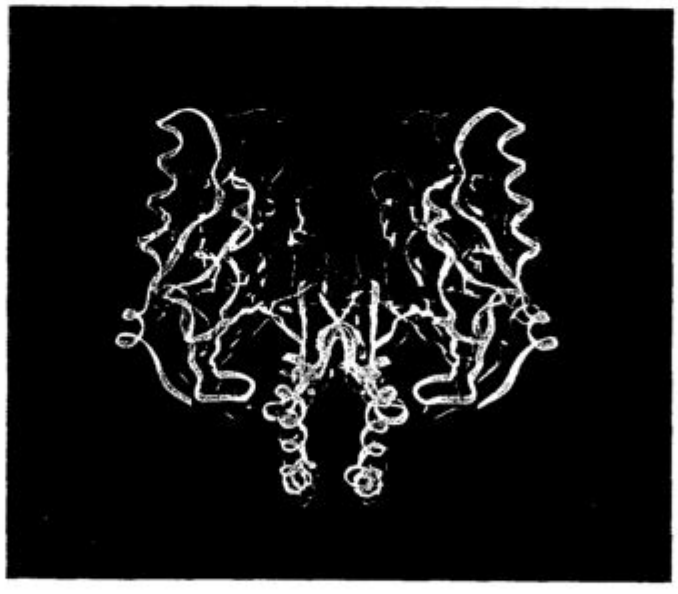
2. Αν η f είναι παραγωγίσιμη στο $x_1=0$ να δειχθεί ότι η f είναι παραγωγίσιμη στο $\mathbb{R}$ με

$$f'(x)=f'(0)\cdot\sigma\upsilon\nu x.$$

α)

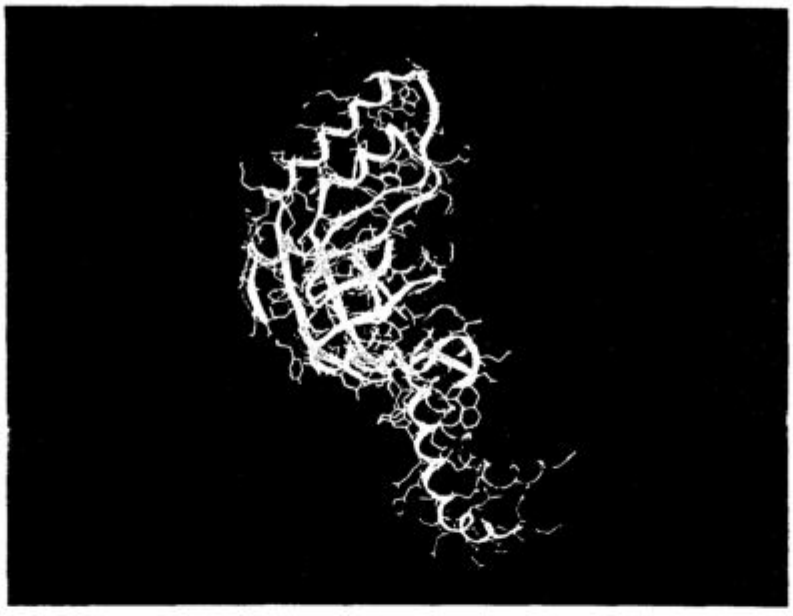


β)

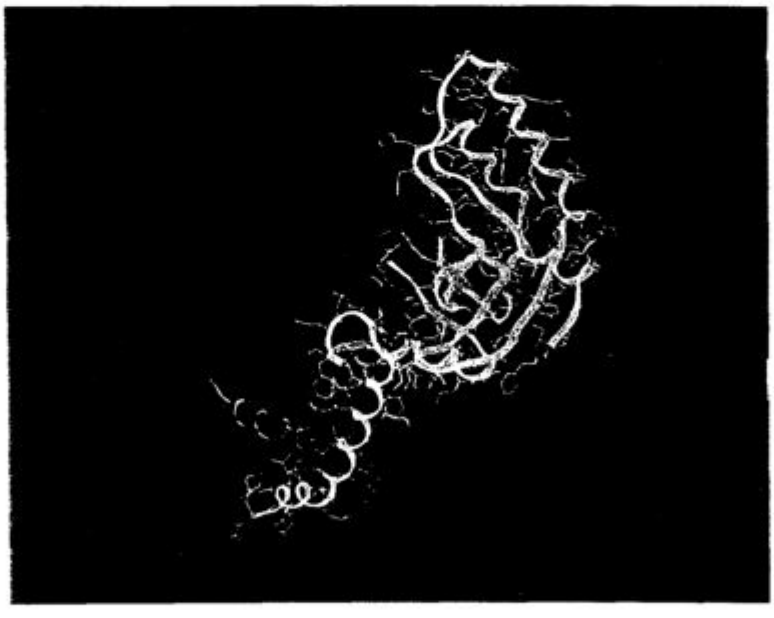


Εικόνα 55. Υπέρθεση της δομής του συμπλόκου R.PvuII-DNA (κόκκινο) με αυτήν της ελεύθερης πρωτείνης (κίτρινο) σε επίπεδο διμερούς. *Για την στοίχιση χρησιμοποιήθηκε η περιοχή διμερισμού (κατάλοιπα 4-45) α) Ο άξονας του DNA (από το μοντέλο του συμπλόκου) στο επίπεδο της φωτογραφίας και β) κάθετα σε αυτό Πάνω στο μοντέλο του συμπλόκου με πράσινο χρώμα οι περιοχές με κατάλοιπα γλυκινης και τέλος με μπλέ χρώμα το μόριο του DNA.*

α)



β)



Εικόνα 56. Υπέρθεση των μονομερών του συμπλόκου με συγγενές DNA (κόκκινο) - ελεύθερης πρωτείνης (κίτρινο). *Για την στοίχιση χρησιμοποιήθηκε τμήμα της C-τελικής υποπεριοχής, χαρακτηρισμένη στην δομή του συμπλόκου σαν περιοχή πρόσδεσης του DNA (κατάλοιπα 58-157) α) μονομερές A και β) μονομερές B Και εδώ με πράσινο χρώμα στο μοντέλο του συμπλόκου τα κατάλοιπα γλυκίνης του μορίου.*

Γ. Συμπεράσματα και προοπτικές

Στα πλαίσια της παρούσης διατριβής προσδιορίσαμε την τρισδιάστατη δομή της ελεύθερης *Pvu*II ενδονουκλεάσης σε ατομική λεπτομέρια. Δείξαμε ότι η R.*Pvu*II είναι ένα διμερές μόριο μεγέθους 156 αμινοξικών καταλοίπων της οποίας ο δομικός πυρήνας δημιουργείται από μία στρεβλωμένη β-πτυχωτή επιφάνεια. Ο διμερισμός του ενζύμου βασίζεται στην δημιουργία δύο δεματιών α-ελικών με εκτεταμένες υδρόφοβες αλληλεπιδράσεις μεταξύ τμημάτων μίας αμφιπαθηκής και κεκαμμένης α-έλικας μεγέθους 5 στροφών και μίας δεύτερης, μεγέθους 3 στροφών, για κάθε μονομερές. Η υδρόφοβη επιφάνεια δημιουργείται κυρίως από κατάλοιπα λευκίνης και ισολευκίνης. Εκτεταμένοι βρόγχοι που συνδέουν στοιχεία της δευτεροταγούς δομής εμφανίζουν ιδιαίτερη ευκινησία στην δομή της ελεύθερης πρωτεΐνης. Η δομή του διμερούς της ελεύθερης R.*Pvu*II εμφανίζει ένα άνοιγμα μεγέθους 25Å στο οποίο ταιριάζει ικανοποιητικά μοντέλο B-μορφής DNA και σε διεύθυνση τέτοια ώστε ο άξονας διμερισμού της πρωτεΐνης να συμπίπτει με τον άξονα συμμετρίας της παλίνδρομης αλληλουχίας του συγγενούς DNA.

Συγκρίνοντας τις δομές των R.*Pvu*II και R.*Eco*RV διαπιστώθηκαν εκτεταμένες τοπολογικές αντιστοιχίες των στοιχείων της δευτεροταγούς δομής τους στην υποπεριοχή αναγνώρισης, πρόσδεσης και κατάλυσης των δύο μορίων. Αντίθετα οι υποπεριοχές διμερισμού των δύο ενζύμων δεν εμφανίζουν δομική συσχέτιση. Την τοπολογική ομοιότητα των δομών των δύο ενζύμων, που δεν σχετίζονται σε επίπεδο αμινοξικής αλληλουχίας, αποδώσαμε στην κοινή ιδιότητα τους να διασπούν φωσφοδιεστερικούς δεσμούς στο κέντρο της αλληλουχίας αναγνώρισης, έχοντας ως αποτέλεσμα το DNA προϊόν να έχει "τυφλά" άκρα. Ενισχύοντας την πρόταση αυτή, η δομή της R.*Bam*HI δείχτηκε {Newman et al., 1994} να εμφανίζει ανάλογης φύσης ομοιότητες με αυτήν της R.*Eco*RI. Τα δύο ένζυμα αυτά πέπτουν το μόριο του DNA αφήνοντας 5' προεξέχοντα άκρα οδηγώντας τους συγγραφείς σε ανάλογα συμπεράσματα για την συσχέτιση των περιοριστικών ενζύμων αυτής της κατηγορίας. Οι δομικές εργασίες πάνω στην R.*Pvu*II και R.*Bam*HI για πρώτη φορά επιτρέπουν την ομαδοποίηση των περιοριστικών ενζύμων με βάση δομικά τους χαρακτηριστικά παρέχοντας παράλληλα και τα πρώτα στοιχεία για την μελέτη της εξελικτικής τους σχέσης. Ο προσδιορισμός δομών περιοριστικών ενζύμων που πέπτουν το μόριο του DNA δημιουργώντας 3' προεξέχοντα άκρα μπορεί να αποτελέσει επιβεβαίωση των προτεινόμενων με βάση τις μέχρι σήμερα προσδιορισμένες δομές. Η ταυτοποίηση δομικά συντηρημένων στοιχείων των περιοριστικών ενζύμων είναι ιδιαίτερης σημασίας για την κατανόηση του μηχανισμού αναγνώρισης και κατάλυσης των ενδονουκλεασών καθώς παρέχει την δυνατότητα αξιολόγισης της σημασίας των καταλοίπων σε πειράματα μεταλαξογέννεσης.

Πέρα από τις προαναφερθείσες ομοιότητες των R.*Pvu*II και R.*Eco*RV παρατηρήθηκαν διαφοροποιήσεις που αποκτούν σημασία για την κατανόηση της λειτουργίας των ενζύμων αυτών. Στην R.*Eco*RV έχουν διαπιστωθεί μικρές αλλαγές στην διαμόρφωση της πρωτεΐνης ανάμεσα στην ελεύθερη μορφή της και στην προσδεδεμένη στην συγγενή αλληλουχία του DNA {Winkler et al., 1993}. Η διαφοροποιήσεις αυτές μπορούν να περιγραφούν ως μετατοπίσεις της υποπεριοχής πρόσδεσης του DNA σε σχέση με την υποπεριοχή διμερισμού του μορίου. Στην R.*Pvu*II οι διαφοροποιήσεις στην διαμόρφωση της πρωτεΐνης ελεύθερης και προσδεδεμένης στο μόριο του DNA αποκτούν δραματικό χαρακτήρα καθώς οι δύο περιοχές πρόσδεσης πλησιάζουν κατά 20Å περίπου. Λεπτομερέστερη ανάλυση των διαφορών διαμόρφωσης της ελεύθερης R.*Pvu*II και της δεσμευμένης σε συγγενή αλληλουχία DNA έδειξε ότι αυτές μπορούν να περιγραφούν ως περιστροφές των υποπεριοχών πρόσδεσης του DNA, και των δύο μονομερών γύρω από μία περιοχή άρθρωσης που περιλαμβάνει τα κατάλοιπα Gly36, Gly37 και Lys38 ανάμεσα στις δύο υποπεριοχές της πρωτεΐνης. Η συμμετοχή καταλοίπων Gly στην περιοχή άρθρωσης είναι μία προφανής εξελικτική επιλογή λόγω του μεγάλου βαθμού στερεοχημικής ελευθερίας που εμφανίζουν. Τα αμινοξέα που μετέχουν στον σχηματισμό της καταλυτικής περιοχής, όπως αυτή ορίζεται τόσο με βάση την ομοιότητα με την R.*Eco*RV έμεσα, όσο και από την δομή του συμπλόκου άμεσα {Cheng et al., 1994}, ανήκουν στην υποπεριοχή πρόσδεσης του DNA. Προς χάριν περιγραφής μπορούμε να πούμε ότι η καταλυτική υποπεριοχή αντιπροσωπεύει πρόσφυση στην περιοχή αναγνώρισης και πρόσδεσης του DNA. Σαν αποτέλεσμα η καταλυτική περιοχή παρασύρεται στις μετακινήσεις της υποπεριοχής πρόσδεσης του DNA. Ετσι αν και ήδη το καταλυτικό κέντρο είναι σχηματισμένο στην ελεύθερη πρωτεΐνη, βρίσκεται τοποθετημένο έκτοπα και σε μεγάλη απόσταση από την περιοχή στην οποία θα επιτελέσει την δράση του. Η περιστροφή των υποπεριοχών πρόσδεσης παρασύρει το καταλυτικό κέντρο οδηγώντας το, στην τελική του θέση. Οι δύο υποπεριοχές πρόσδεσης του DNA (μια για κάθε μονομερές) στην ελεύθερη R.*Pvu*II βρίσκονται σε τέτοια απόσταση που δεν επιτρέπει (βάσει μοντέλου B-DNA) την ταυτόχρονη αλληλεπίδραση με το DNA, πλευρικών αλυσιδών των καταλοίπων τους, που με βάση την δομή του συμπλόκου μετέχουν στις ειδικές αλληλεπιδρασεις αναγνώρισης της αλληλουχίας. Αντίθετα η δομή του συμπλόκου, με το συγγενές DNA, εμφανίζει μια ιδιαίτερα κλειστή δομή, αγκαλιάζοντας το μόριο του DNA με τρόπο που καθιστά απίθανη την αναζήτηση της συγγενούς αλληλουχίας από αυτήν την διαμόρφωση της πρωτεΐνης. Με βάση τα προηγούμενα, μπορούν να γίνουν δύο υποθέσεις, για την διαμόρφωση της πρωτεΐνης στο στάδιο της αναζήτησης της συγγενούς αλληλουχίας. Η πρώτη υποθέτει την ύπαρξη μιάς τρίτης διαμόρφωσης της πρωτεΐνης προσδεδεμένης σε DNA μή συγγενούς

αλληλουχίας. Μέχρι σήμερα δεν έχει δειχθεί κάτι τέτοιο καθώς η προσπάθεια για την απομόνωση και κρυστάλλωση ενδονουκλεασών σε σύμπλοκο με τη μή συγγενή αλληλουχία υπήρξαν ανεπιτυχείς. Στην μόνη περίπτωση όπου ολιγονουκλεοτίδιο με μή συγγενή αλληλουχία συγκρυσταλλώθηκε με ενδονουκλεάση, η επίλυση της δομής έδειξε ότι η πρόσδεση έγινε με μή τυπικό τρόπο {Winkler et al., 1993}. Στην δεύτερη υπόθεση, την οποία θεωρούμε ότι υποστηρίζουν τα δεδομένα της παρούσης εργασίας, η ανίχνευση της αλληλουχίας πραγματοποιείται στην διαμόρφωση της πρωτεΐνης στην ελεύθερη μορφή της, από κατάλοιπα στην περιοχή διέλευσης του άξονα διμερισμού του μορίου και ταυτόχρονα στην περιοχή άρθρωσης μεταξύ των δύο υποπεριοχών της πρωτεΐνης. Στην περίπτωση αυτή η ανίχνευση αφορά το κεντρικό ζεύγος βάσεων της αλληλουχίας αναγνώρισης GC. Η ικανοποίηση αλληλεπιδράσεων στην περιοχή αυτή οδηγεί στις παρατηρούμενες αλλαγές στην διαμόρφωση και τον έλεγχο των υπολοίπων βάσεων της αλληλουχίας αναγνώρισης. Η δεύτερη αυτή υπόθεση υποστηρίζεται από μία σειρά δεδομένων από την μοριακή, βιοχημική και δομική αναλύση τόσο της R.*Pvu*II όσο και των υπολοίπων ενδονουκλεασών : Παρά την μεγάλη ειδικότητα που εμφανίζουν οι ενδονουκλεάσες σε ότι αφορά την αλληλουχία αναγνώρισης του DNA, στην δομή του συμπλόκου της R.*Eco*RV δεν εμφανίζονται ειδικές αλληλεπιδράσεις με τις δύο κεντρικές βάσεις της συγγενούς αλληλουχίας {Winkler et al., 1993} ενώ στην δομή της R.*Pvu*II (σύμπλοκο) εμφανίζεται ένας πιθανός υδρογονικός δεσμός με την μεσολάβηση μορίου νερού αμφίβολης συμβολής στην ειδικότητα της αναγνώρισης {Cheng et al., 1994}. Στο σύμπλοκο της R.*Pvu*II με το DNA τα γειτονικά προς την περιοχή άρθρωσης (Gly36-Lys38) κατάλοιπα Gln33-Asp34-Asn35 μετέχουν σε ειδικές αλληλεπιδράσεις στο μικρό αυλάκι του DNA {Cheng et al., 1994}. Στην ελεύθερη πρωτεΐνη το Gln33 κατάλοιπο του μονομερούς A αλληλεπιδρά με το Asn35 του μονομερούς B ενώ η Lys38 με το κατάλοιπο Asp30 στην αρχή του βρόγχου LAB δημιουργώντας ένα πλέγμα αλληλεπιδράσεων που ενδεχόμενα έχει συμμετοχή στην σταθεροποίηση της σχετικής διεύθυνσης των υποπεριοχών πρόσδεσης του DNA. Εναλλακτική ικανοποίηση των αλληλεπιδράσεων των καταλοίπων του βρόγχου LAB με αυτές που παρουσιάζονται στην δομή του συμπλόκου μπορεί να αποτελέσει έναν απο τους παράγοντες που σχετίζονται με τις αλλαγές διαμόρφωσης του μορίου. Μεταλλαξογέννεση της R.*Eco*RV, καταλοίπων που πραγματοποιούν ειδικές αλληλεπιδράσεις με βάσεις του DNA έχει το παράδοξο αποτέλεσμα: να αφήνουν ανεπηρέαστη την ικανότητα του ενζύμου να αναγνωρίζει και να προσδένεται στην ειδική αλληλουχία του DNA, μειώνοντας όμως, σημαντικά την καταλυτική ενεργότητα του {Vermote, Vipond and Halford, 1992}. Τα αποτελέσματα αυτά μπορούν ερμηνευτούν κάτω απο το πρίσμα ότι αλληλεπιδράσεις πλευρικών αλυσίδων της πρωτεΐνης με βάσεις του DNA δεν συμμετέχουν απαραίτητα στην

ειδικότητα του ενζύμου αλλά σχετίζονται με την αναγκαιότητα της ακριβούς τοποθέτησης της πρωτεΐνης πάνω στο DNA για την δημιουργία του καταλυτικού κέντρου. Αυτό προϋποθέτει ότι μέρος τουλάχιστον των αλληλεπιδράσεων αναγνώρισης έχει προηγηθεί. Επιπλέον έχει δειχθεί ότι η R.*Eco*RV απουσία Mg^{2+} δεν έχει προτίμηση για την συγγενή της αλληλουχία {Thielking et al., 1992}. Η περιοχή όπου πραγματοποιούνται οι ειδικές αλληλεπιδράσεις στο σύμπλοκο βρίσκεται σε μεγάλη απόσταση απο την περιοχή πρόσδεσης των ιόντων Mg^{2+}. Με βάση αυτά τα δεδομένα έχει υποτεθεί η ύπαρξη επικοινωνίας μεταξύ των λειτουργιών της αναγνώρισης με αυτή της πρόσδεσης του Mg^{2+} χωρίς να έχει αποσαφηνιστεί ο μηχανισμός της {Vermote & Halford, 1992}. Αντίθετα η περιοχή του βρόγχου Q που είναι τοπολογικά ισόδύναμη με αυτήν του LAB βρόγχου της R.*Pvu*II βρίσκεται σε γειτονική θέση με αυτήν της πρόσδεσης του Mg^{2+} και αποτελεί πιθανότερο υποψήφιο για την επικοινωνία μεταξύ αναγνώρισης και κατάλυσης. Τέλος αντικατάσταση καθεμίας βάσης T απο της δύο του DNA-στόχου της EcoRI μεθυλάσης έχει σαν αποτέλεσμα την πτώση της σταθεράς πρόσδεσης της πρωτείνης στο DNA όπως αναμένεται λόγω της μείωσης του αριθμού υδρογονοδεσμών μεταξύ πρωτείνης-DNA. Ομως αντικατάσταση και των δύο βάσεων T δεν εμφανίζει το αναμενόμενο προσθετικό αποτέλεσμα στην σταθερά πρόσδεσης της πρωτείνης {Reich & Danzitz, 1991}. Ανάλογα είναι τα αποτελέσματα απο αντικατάσταση βάσεων G με Ινοσίνη (I) {Reich & Danzitz, 1992}. Τα αποτελέσματα αυτά ερμεινεύονται από ένα μοντέλο σταδιακής ειδικής αναγνώρισης όπου κάθε μία από τις βάσεις που μεταλλάσονται ανήκει σε διαφορετικό στάδιο της διαδιακασίας αναγνώρισης της αλληλουχίας. Ενας τέτοιος μηχανισμός μπορεί να έχει εφαρμογή σε πρωτείνες με ενζυματική δράση σε μόρια DNA καθώς ελαχιστοποιεί την πιθανότητα της καταλυτικής λειτουργίας σε λάθος αλληλουχίες,

Στην δομή του συμπλόκου της R.*Pvu*II το DNA βρίσκεται σε περίπου ιδανική B-μορφή {Cheng et al., 1994}, αυτό έρχεται σε πλήρη αντίθεση με τα αποτελέσματα για τις R.*Eco*RI και R.*Eco*RV όπου το DNA υπόκειται σε σημαντικές αποκλίσεις απο την B μορφή. Από την σύγκριση της R.*Pvu*II με την R.*Eco*RV (που μπορεί να γινει άμεσα λόγω της τοπολογικής ομοιότητας, εικόνες 51, 52) βλέπουμε ότι η R.*Pvu*II στερείται των εκτεταμένων βρόγχων της R.*Eco*RV που εισχωρούν στο μεγάλο αυλάκι του DNA. Στην R.*Pvu*II τα κατάλοιπα που μετέχουν στις ειδικές αλληλεπιδράσεις έχουν περιορισμένη συγκριτικά ευκινησία σε σχέση με τα αντίστοιχα της R.*Eco*RV. Τα δύο ένζυμα φαίνεται να ακολουθούν εν μέρει διαφορετικούς μηχανισμούς επαγώμενης προσαρμογής (induced fit) κατά την πρόσδεση. Στην R.*Eco*RV οι αλλαγές στην διαμόρφωση της πρωτεΐνης, αν και υπαρκτές, είναι μικρές σε έκταση και είναι το μόριο του DNA πού υφίσταται τις συνέπειες της πρόσδεσης, αντίθετα στην R.*Pvu*II το σύνολο των αλλαγών

διαμόρφωσης αφορά το πρωτεϊνικό μόριο. Τα τελευταία χρόνια ο μηχανισμός της επαγώμενης προσαρμογής κατά την ειδική πρόσδεση πρωτεϊνών στο DNA αποδεικνείεται να έχει εφαρμογή σε ολοένα και περισσότερα δομικά πρότυπα πρωτεϊνών που προσδένονται σε νουκλεικά οξέα σε αντίθεση με την αντίληψη που ήθελε την αναγνώριση του DNA να είναι του τύπου κλειδιού - κλειδαριάς. Η κατανόηση των ενεργειακών παραμέτρων της διαδικασίας επαγώμενης προσαρμογής μόλις αρχίζει να γίνεται κατανοήτη { Spolar & Record, 1994}.

Η αδυναμία για την πλήρη κατανόηση μέχρι την στιγμή αυτή του μηχανισμού αναγνώρισης της ειδικής αλληλουχίας του DNA από τα περιοριστικά ένζυμα, σχετίζεται με την αποτυχία απομόνωσης των ενδιάμεσων καταστάσεων μεταξύ της ελεύθερης μορφής και αυτής του συμπλόκου με την συγγενή αλληλουχία. Στην περίπτωση της R.*Pvu*II δεν έχουν πραγματοποιηθεί προσπάθειες στην κατεύθυνση αυτή. Η R.*Pvu*II είναι το δεύτερο ένζυμο, της κατηγορίας του, για το οποίο είναι γνωστές οι δύο ακραίες διαμορφώσεις του, έτσι η επιτυχία πειράματων συγκρυστάλλωσης με μή συγγενή αλληλουχία και η δομική ανάλυση τους μπορεί να δώση πληροφορία άμεσα αξιοποιήσιμη. Επιπλέον η δομική ανάλυση συγκρυστάλλων με την *Pvu*II* αλληλουχία αναγνώρισης, θα επιτρέψει την άμεση σύγκριση των ειδικών αλληλεπιδράσεων που διαφοροποιούν την αναγνώριση των δύο αλληλουχιών.

Μεταλλαξογέννεση των ευκίνητων καταλοίπων, που μετέχουν στην άρθρωση του μορίου, Gly36 και Gly37 σε κατάλοιπα με μικρότερη ευελιξία (Ala) ή κατάλοιπα που δημιουργούν στερεοχημική παρεμπόδιση (Pro) σε αντιπαράθεση με μεταλλαγές της Lys38 με κατάλοιπα είτε με διαφορετική ενεργή ομάδα (π.χ. Glu) είτε με μικρότερη πλευρική αλυσίδα (π.χ. Asn) μπορεί να δώσει απάντηση στο ερώτημα κατα πόσον ο ρόλος της περιοχής είναι παθητικός και απλά δέχεται τις αλλάγες διαμόρφωσης που οφείλονται σε δυνάμεις που αναπτύσονται σε διαφορετικές περιοχές του μορίου, η αν το κατάλοιπο Lys38 μετέχει σε κάποιο μηχανισμό επαγωγής των αλλαγών διαμόρφωσης.

Η Pro14 στο μέσον της α-έλικας αA της R.*Pvu*II έχει ως αποτέλεσμα μία κάμψη 30° περίπου ανάμεσα στα δύο τμήματα της. Η κάμψη της έλικας αA είναι σημαντική για τον καθορισμό της σχετικής θέσης των δύο μονομερών. Είναι ενδιαφέρον ότι οι τρείς (R.*Pvu*II, R.*Bam*HI {Newman et al., 1994} και R.*Eco*RV {Winkler et al., 1993}) από τις τέσσερεις δομές περιοριστικών ενζύμων, που έχουν επιλυθεί περιέχουν μία κεκκαμένη έλικα λόγω παρουσίας καταλοίπου Pro στο εσωτερικό της. Οι κεκκαμένες αυτές έλικες εμφανίζονται στις τρείς δομές σε περιοχές που εκπληρώνουν διαφορετικούς ρόλους. Είναι γνωστό ότι η παρουσία καταλοίπων Pro στο εσωτερικό α-ελίκων είναι ιδιαίτερα σπάνια {MacArthur & Thornton, 1991}, {Barlow & Thornton, 1988}. Με δεδομένο ότι οι αλληλουχίες των ενζύμων αυτών δεν σχετίζονται μεταξύ τους, μπορεί να υποτεθεί ότι είτε η

Pro στο εσωτερικό των α-ελικών αποτελεί ένα ιδιαίτερα συντηρημένο κατάλοιπο από μια πολύ απομακρυσμένη εξελικτικά προγονική πρωτεΐνη, είτε ότι αυτές επιτελούν κάποιο ιδιαίτερο και αδιευκρίνιστο δομικό ρόλο στα περιοριστικά ένζυμα. Μεταλλαγή της Pro14 σε κατάλοιπο που να καταργεί την κάμψη της αΑ έλικας χωρίς να αλλοιώνει το πρότυπο αμφιπαθηκότητας της μπορεί να δώσει την απαραίτητη πληροφορία για την κατανόηση του δομικού ρόλου της.

Δ. Υλικα και Μέθοδοι

Δ 1. Υλικά

Στα στάδια καθαρισμού της R.*Pvu*II χρησιμοποιήθηκαν τα χρωματογραφικά υλικά : Q-FastFlow, MonoQ (προπακεταρισμένη στήλη), Sephadex 75, Superdex 75 (προπακεταρισμένη στήλη) τα οποία αγοράστηκαν από την εταιρεία Pharmacia. Από την εταιρεία Whatman χρησιμοποιήθηκε η φωσφοκυτταρίνη τύπου P11 ύστερα από επεξεργασία που αναφέρεται στην παράγραφο Δ 2.7.

Για την παρασκευή των διαλυμάτων έκλουσης των κολωνών χρησιμοποιήθηκαν KCl, K_2PO_4 και Trizma Base της εταιρείας Sigma, K_2HPO_4 και KH_2PO_4 της εταιρείας Merck. Το Tris-HCl Ultrapure ήταν της εταιρείας BRL. Σαν πρόσθετα χρησιμοποιήθηκαν 2-Mercaptoethanol της Merck και Dithiothreitol (DTT) της Sigma. Χρησιμοποιήθηκαν αναστολείς της πρωτεόλυσης Banzamidine, Leupeptin (Hemisulfate Salt) και PMSF (Phenyl Methyl Sulfonyl Fluoride) της εταιρείας Sigma.

Στην εκχύλιση των βακτηριακών κυττάρων χρησιμοποιήθηκε Λυσοζύμη (Chiken egg white) της Sigma και η υδρόλυση νουκλεικών οξέων του βακτηριακού εκχυλίσματος έγινε με DNase I (Bovine Pancreas) της ίδιας εταιρείας.

Για συμπυκνώσεις της πρωτεΐνης και αλλαγή ρυθμιστικού διαλύματος χρησιμοποιήθηκαν μεμβράνες της εταιρείας Spectrapor με κατώφλι μοριακού βάρους 6000-8000Da. Φιλτράρισμα των πρωτεϊνικών δειγμάτων έγινε με φίλτρα 0.22 και 0.45μm της Millipore. Για την συμπύκνωση της πρωτεΐνης χρησιμοποιήθηκε PEG20000 της εταιρείας Merck. Σε συμπυκνώσεις της πρωτεΐνης με την μέθοδο της υπερδιήθησης χρησιμοποιήθηκαν τα Centricon 10 και Centriprep 10 της εταιρείας Amicon.

Για τις κρυσταλλώσεις χρησιμοποιήθηκε θειικό αμμώνιο μεγάλης καθαρότητας της Aldrich ενώ για κατακρήμνιση πρωτεϊνών στα στάδια του καθαρισμού το θειικό αμμώνιο ήταν της εταιρείας Sigma. Τέλος για κρυσταλλώσεις επίσης χρησιμοποιήθηκαν PEG διαφορετικών μοριακών βαρών των εταιρειών Serva και Merck. Ρυθμιστικά διαλύματα για τις κρυσταλλώσεις έγιναν με οξικό αμμώνιο της εταιρείας Sigma. Επίσης ως παράγοντας κατακρήμνισης χρησιμοποιήθηκε Lithium Sulfate τηε εταιρείας Sigma.

Πηκτώματα αγαρόζης παρασκευάστηκαν με Ultrapure αγαρόζη της εταιρείας BRL. Για μάρτυρες σε πηκτώματα ακρυλαμίδης χρησιμοποιήθηκε το LMW Electrophoresis Calibration Kit της εταιρείας Pharmacia.

Συνθετικά ολιγονουκλεοτίδια κατασκευάστηκαν από το εργαστήριο Μικροχημείας του IMBB ενώ για τον προσδιορισμό της νουκλεοτιδικής αλληλουχίας χρησιμοποιήθηκε το προϊόν Sequenase 2.0 της εταιρείας USB (United

States Biochemicals). Τα ράδιοσημασμένα νουκλεοτίδια προήλθαν από την εταιρεία Amersham. Στην ανάλυση του κλώνου της R.*Pvu*II χρησιμοποιήθηκαν περιοριστικά ένζυμα των εταιρειών New England Biolabs και Minotech. Το DNA αδενοιού 2 για τα πειράματα υπερπέψης ήταν επίσης της εταιρείας Minotech.

Για την πραγματοποίηση κρυσταλλογραφικών μετρήσεων οι κρύσταλλοι εγκλείστηκαν σε γυάλινα τριχοειδή σωληνάρια (glass capillaries) της εταιρείας Mark-Rohrchen διαμετρήματος 0.3, 0.5, 0.7, 1.0mm.

Δ 2. Απομόνωση και ανάλυση νουκλεικών οξεών.

Δ 2.1. Απομόνωση πλασμιδιακού DNA.

Απομόνωση πλασμιδιακού DNA σε μικρή κλιμακα πραγματοποιήθηκε με την μέθοδο βρασμού {Holmes & Quingley, 1981} και σε μεγάλη κλίμακα (500ml) πραγματοποιήθηκε με τροποποίηση της μέθοδου αλκαλικής λύσης {Birnboim, 1983}, {Sambrook, Fritch and Maniatis, 1989} και απομόνωση της πλασμιδιακής ζώνης από διαβάθμιση CsCl {Radloff, Bauer and Vinograd, 1967}.

Δ 2.2. Ανάλυση τμημάτων DNA σε πήκτωμα αγαρόζης.

Ανάλυση των προϊόντων πέψης DNA κατά τον προσδιορισμό της (*) αλληλουχίας αναγνώρισης της R.*Pvu*II καθώς και για την μέτρηση της ενζυμικής δραστικότητας της R.*Pvu*II στα στάδια καθαρισμού πραγματοποιήθηκε σε πηκτώματα 1% αγαρόζης {McDonell, Simon and Studier, 1977}. Χρώση των ζωνών του DNA στα πηκτωμάτα έγινε με βρωμιούχο αιθίδιο {Sharp, Sugden and Sambrook, 1973}.

Δ 2.3. Προσδιορισμός νουκλεοτιδικής αλληλουχίας μορίων DNA .

Για τον προσδιορισμό της νουκλεοτιδικής αλληλουχίας ακολουθήθηκε η μέθοδος που βασίζεται στον τερματισμό της επιμήκυνσης της DNA αλυσίδας λόγω της ενσωμάτωσης δι-δεόξυνουκλεοτιδίων {Sanger, Nicklen and Coulson, 1977}. Το μεγαλύτερο μέρος του προσδιορισμού έγινε με την χρήση τροποποιημένης T7 πολυμεράσης με την επονομασία Sequenase. Σε μερικές περιπτώσεις χρησιμοποιήθηκε το μεγάλο τμήμα της DNA πολυμεράσης I (Klenow fragment).

Ακολουθήθηκε το πρωτόκολλο που περιγράφεται στο συνοδευτικό εγχειρίδιο του κιτ, για προσδιορισμό της αλληλουχίας του DNA {USB, 1990} Sequenase 2.0, της USB (United States Biochemicals).

Στις περιπτώσεις που χρησιμοποιήθηκε το μεγάλο κομμάτι της πολυμεράσης (Klenow fragment) ακολουθήθηκε τροποποίηση της διαδικασίας για τον προσδιορισμό της αλληλουχίας υπερελικομένου DNA {Barlett, Gaillard and Joklik, 1986} :

1) 1-2 μgr πλασμιδιακού DNA πέπτονται με περίσσεια περιοριστικού ενζύμου σε θέση αντίθετη από αυτήν του εκκινητή που θα χρησιμοποιηθεί σε σχέση με τις θέσεις ένθεσης.

2) 100 ng υδρολυμένου DNA αναμιγνύονται με 1.5 μl ρυθμιστικό διάλυμα αντίδρασης RB (10xRB : 500mM Tris-Cl pH7.6, 500 mM KCl, 100 mM $MgCl_2$ και 100 mM DTT), 7 μl εκκινητή και συμπληρώνουμε μέχρι τα 15 μl με nanopure νερό.

3) Ακολουθεί θέρμανση του μίγματος στους 100° C για 3΄ και τοποθέτηση του στον πάγο.

4) Προστίθενται 1 μl α[32P]-dATP και 1μl Klenow (5u/μl).

5) Σε τρυβλίο πολλαπλών θέσεων βάζουμε από 1 μl για κάθε μίγμα δι-δεοξυνουκλεοτιδίου.

6) Από το μίγμα της αντίδρασης 3.3 μl μεταφέρονται σε κάθε ένα από τα 4 πηγάδια.

7) Επώαση για 15΄ σε θερμοκρασία δωματίου.

8) Μετά την επώαση προστίθεται 1μl ρυθμιστικού διαλύματος συμπλήρωσης (chase buffer: 0.25 mM από κάθε dNTP σε RB) σε κάθε αντίδραση και συνεχίζουμε την επώαση για 5΄.

9) Ακολουθεί προσθήκη ρυθμιστικού διαλύματος φόρτωσης (Loading Buffer).

Loading Buffer : 20 μl 0.5M EDTA, 10 μl NaOH 10N, 70 μl νερό
Προσθήκη 400 μl ουρίας 1mg/ml
Προσθήκη κόκκων Xylene-Cyanol, Bromophenol Blue.

10) Στο στάδιο αυτό τα δείγματα ηλεκτροφορούνται σε αποδιατακτικό πήκτωμα ακρυλαμίδης (8M ουρία) {Sanger & Coulson, 1978} η φυλάσσονται στους -20° (για μικρό χρονικό διάστημα).

Δ 3. Παραγωγή, απομόνωση και χαρακτηρισμός της πρωτεΐνης.

Δ 3.1. Μεγάλης κλίμακας καλλιέργεια βακτηρίων για την παραγωγή πρωτεΐνης.

Η παραγωγή της R.*Pvu*II πρωτεΐνης έγινε από κύτταρα του *E.coli* στελέχους HB101. Το στέλεχος HB101 χαρακτηρίζεται απο την απουσία των *McrA, McrB* και *Mrr* συστημάτων περιορισμού των *E.coli* κυττάρων τα οποία

πέπτουν μεθυλιωμένο DNA. Η χρήση του συγκεκριμένου στελέχους είναι ιδιαίτερης σημασίας, καθώς επιτρέπει την παραγωγή της *Pvu*II μεθυλάσης η οποία έχει ως αποτέλεσμα την δημιουργία DNA ευάλωτου στα προηγούμενα συστήματα. Οι καλλιέργιες των βακτηριακών κυττάρων πραγματοποιήθηκαν με καθιερώμενα θρεπτικά μέσα για κύτταρα *E.coli* (Superbroth). Για τις καλλιέργειες ακολουθήθηκαν τα βήματα που παρουσιάζονται πιο κάτω. Κάθε καλλιέργεια απέδωσε κατά μέσον όρο 200gr βακτηριακής πάστας (περίπου 6gr/lt καλλιέργειας).

1) Κύτταρα *E.coli* τα οποία φέρουν το πλασμίδιο p*Pvu*Prm3.4 φυλάσσονται στους -70°C, σε αιώρημα 30% γλυκερόλης. Μικρή ποσότητα κυττάρων επιστρώνεται σε τρυβλία petri με θρεπτικό μέσο LB στερεοποιημένο με άγαρ έχοντας ως πρόσθετο αμπικιλλίνη 50μgr/ml. Επωάζονται στους 37° C για 18 ώρες.

2) Μία βακτηριακή αποικία από τα τρυβλία μεταφέρεται σε 3ml υγρή καλλιέργεια σε θρεπτικό μέσο LB παρουσία αμπικιλλίνης. Ακολουθεί επώαση στους 37° για 12 ώρες.

3) Από 1ml της προηγούμενης καλλιέργειας μεταφέρεται σε τρείς φλάσκες με θρεπτικό μέσο LB και προσθήκη αμπικιλλίνης και ακολουθεί επώαση στους 37° C για 12 ώρες.

4) Σε ζυμωτήρα 30 lt αναμιγνύονται 480gr Τρυπτόνη, 300gr Yeast extract, 150gr NaCl και αποστειρώνονται. Οταν το θρεπτικό μέσο αποκτήσει θερμοκρασία περιβάλλοντος γίνεται προσθήκη αμπικιλλίνης σε τελική συγκέντρωση 35-50 μg/ml.

5) Προσθήκη στο ζυμωτήρα των 3 lt από τις καλλιέργειες στις φλάσκες. Επώαση στους 37° με ανάδευση 100rpm, παροχή αποστειρωμένου αέρα και περιοδικό έλεγχο και ρύθμιση του pH της καλλιέργειας ώστε να διατηρείται στα όρια 7.0-7.5. Η ρύθμιση γίνεται με σταδιακή προσθήκη οξέος (HCl) ή βάσης (NaOH). Τα κύτταρα επωάζονται στον ζυμωτήρα για 12-16 ώρες (τελική OD_{600}=4).

6) Συλλογή των κυττάρων σε μορφή βακτηριακής πάστας με φυγοκεντρική μέθοδο.

7) Αποθήκευση της βακτηριακής πάστας στους -70°C.

Δ 3.2. Λύση πάστας βακτηριακών κυττάρων.

Η υπερηχοβόλιση χρησιμοποιήθηκε για την λύση των κυττάρων στους αρχικούς καθαρισμούς της πρωτεΐνης. Η μέθοδος αν και εξασφαλίζει υψηλή απόδωση ρίξης των κυτταρικών τοιχωμάτων έχει ως μειονέκτημα την αποδιάταξη σημαντικής ποσότητας πρωτεΐνων λόγω τοπικής υπερθέρμανσης του κυτταρικού εναιωρήματος. Η μέθοδος της ενζυματικής λύσης εξασφαλίζοντας ήπιες συνθήκες για το πρωτεϊνικό εκχύλισμα μειώνει τις απώλειες της πρωτεΐνης και κρίθηκε

αποτελεσματικότερη. Οι δύο μέθοδοι έχουν περιγραφεί {Cull & McHenry, 1990} και εφαρμόστηκαν με μικρές τροποποιήσεις.

α) **Υπερηχοβόλιση.** Βακτηριακή πάστα ν gr διαλύεται σε 2ν ml διάλυμα Α (Διάλυμα Α : 10mM Tris pH 7.5, 7mM 2-mercaptethanol). Στο αιώρημα γίνεται προσθήκη αναστολέων πρωτεασών 1mM PMSF, 20μgr/ml Leupeptin, 80 μgr Aprotinin, 150 μgr/ml Benzamidine. Ακολουθεί υπερηχοβόλιση για 12Χ30'' με διαλύματα 30'' (Το δείγμα πρέπει να βρίσκεται σε δοχείο με πάγο) και το εκχύλισμα φυγοκεντρείται για 1 ώρα στα 15000 g. Το υπερκείμενο απομακρύνεται και αποθήκευεται στους 4°. Το ίζημα επαναδιαλύεται σε διάλυμα Α και υπερηχοβολείται για 6Χ30'' με διαλύματα 30'' (Το δείγμα βρίσκεται σε δοχείο με πάγο). Υστερα από νέα φυγοκέντρηση για 30΄ στα 15000 g τα δύο υπερκείμενα ενώνονται και υπερφυγοκεντρούνται για 1 1/2 ώρα σε 40000 g. Το υπερκείμενο υπόκειται σε αλλαγή ρυθμιστικού διαλύματος και φορτώνεται στην πρώτη χρωματογραφική στήλη.

β) **Ενζυματική λύση.** 100-150 gr κυτταρικής πάστας κατεψυγμένης στους -70° C αφήνονται για 12 ώρες στους 4° C. Η κυτταρική πάστα διαλύεται στο ρυθμιστικό διάλυμα λύσης (50mM Tris pH 7.5, 200mM NaCl, 5% γλυκερόλη κατ' όγκο, 1mM DTT και 1mM phenylmethanesulfonyl fluoride (PMSF)) σε όγκο περίπου ίσο αριθμητικά με τη μάζα της. Ακολουθεί προσθήκη λυσοζύμης (300 μg για κάθε ml του εναιωρήματος των κυττάρων) και επώαση στους 4° C για μία ώρα (με ήπια ανάδευση). Στην διάρκεια της ωριαίας επώασης με λυσοζύμη γίνεται προσθήκη αναστολέων των πρωτεασών τμηματικά (π.χ. κάθε 15΄). (Συνολικό μίγμα αναστολέων : 1mM PMSF, 20μgr/ml Leupeptin, 80 μgr/ml Aprotinin, 150 μgr/ml Benzamidine) Στην συνέχεια πραγματοποιείται προσθήκη α) ποσότητας $MgCl_2$ ώστε η συγκέντρωση του διαλύματος να γίνει 10mM και β) 10 μg DNAse I για κάθε ml του αιωρήματος των κυττάρων. Το μίγμα επώαζεται στους 4° C για 30΄ (με ήπια ανάδευση) και στην συνέχεια φυγοκεντρείται για μία ώρα σε 15000g για την απομάκρυνση των κυτταρικών υπολειμάτων. Το υπερκείμενο είναι έτοιμο για μεταφορά στο διάλυμα φορτώματος της πρώτης στήλης με διαπίδιση μεμβράνης.

Δ 3.3. Προετοιμασία υλικού για την κατασκευή στήλης φωσφοκυτταρίνης (PC).

Για την κατασκεύη της στήλης φωσφοκυτταρίνης χρησιμοποιήθηκε το υλικό Phosphocellulose Whatman P11 σε μορφή σκόνης. Το υλικό απαιτεί την προετοιμασία που περιγράφεται προκειμένου να έρθει σε μορφή κατάλληλη για το πακετάρισμα της στήλης. Ολα τα στάδια που περιγράφονται στην συνέχεια

ακολουθούνται από ήπια ανάδευση και αναμονή μέχρι την κατακάθιση. Το υλικό στην όλη διαδικασία δεν πρέπει να παραμείνει σε pH >10 ή pH < 3 για μεγάλα χρονικά διαστήματα. Μέσα σε δοχείο 5lt προστίθενται 150gr PC και 4lt NaOH 0.1N ακολουθούν 3 αλλαγές με νερό nonopure 4lt καθεμία. Στην συνέχεια γίνεται προσθήκη 4lt HCl 0.1N και εκ νέου 3 αλλαγές με νερό nanopure 4lt καθεμία. Τέλος προστίθενται 4lt PO_4^{2-}0.1M με pH 7.0 και πραγματοποιούνται τρείς αλλαγές με το ίδιο διάλυμα. Το υλικό παραμένει στα PO_4^{2-} για 12 ώρες πρίν χρησιμοποιηθεί. Αν χρειαστεί το υλικό φυλάσεται σε PO_4^{2-} 0.1 M pH 7.0, προαιρετική είναι η προσθήκη 0.01% αζιδίου του νατρίου για την αποφυγή ανάπτυξης μικροοργανισμών.

Δ 3.4. Μέθοδος καθαρισμού της *Pvu*II ενδονουκλεάσης.

Εκχύλισμα βακτηριακών κυττάρων από ενζυματική λύση (βλ. κεφάλαιο Δ 3.2.β) μεταφέρεται σε μεμβράνες διαπίδισης και αφήνεται να εξισορροπηθεί σε 2 lt διαλύματος PC : (20mM φωσφορικού καλίου pH7.4, 1mM EDTA, και 10mM 2-μερκαπταιθανόλη) με 100mM KCl για δώδεκα ώρες με μία ενδιάμεση αλλαγή διαλύματος PC ακολουθεί φυγοκέντρηση για την απομάκρυνση πρωτεϊνών που καθιζάνουν στις συνθήκες αυτές.

Το υλικό φορτώνεται σε στήλη φωσφοκυτταρίνης όγκου 300-400ml εξισσοροπημένη με διάλυμα PC. Ακολουθεί πλύσιμο της στήλης με το ίδιο διάλυμα και έκλουση των πρωτεϊνών με 2 lt γραμμικής διαβάθμισης της συγκέντρωσης KCl 0.1M-0.6M. Η R.*Pvu*II εκλούεται περίπου σε 300mM KCl.

Τα καταλυτικά ενεργά κλάσματα (βλ. κεφάλαιο Δ 3.5.) από την στήλη φωσφοκυτταρίνης συλλέγονται και εξισορροπούνται σε μεμβράνη διαπίδισης με διάλυμα Q: (20mM Tris.HCl pH 7.4, 1mM EDTA, 10mM 2-μερκαπτοαιθανόλη) που περιέχει 10mM KCl και ακολουθεί εκ νέου φυγοκέντριση στις 10000 crm για την απομάκρυνση των πρωτεϊνών που καθιζάνουν. Το δείγμα φορτώνεται σε στήλη MonoQ 8ml ή σε στήλη Q-Sepharose Fast Flow 50ml και ακολουθεί εκτεταμένη πλύση (20 X τον όγκο της στήλης). Συλλέγονται κλάσματα από διαβάθμιση συγκέντρωσης KCl 0.01M-0.3M. Τα ενεργά κλάσματα συγκεντρώνονται σε όγκο 1-5 ml και φωρτώνονται σε στήλη μοριακής διήθησης Sephadex75 ή Superdex75 χωρίς άλλη εξισορρόπηση. Τα ενεργά κλάσματα από την στήλη μοριακής διήθησης εξισορροπούνται στο διάλυμα αποθήκευσης: 20mM Tris.HCl pH 7.4 και 20mM KCl και στην συνέχεια συγκεντρώνονται στον επιθυμητό όγκο με υπερδιήθηση. Αποθήκευση της πρωτεΐνης γίνεται στους 4° C και για μικρό χρονικό διάστημα για την αποφυγή του φαινομένου της συσσωμάτωσης (aggregation).

Δ 3.5. Αντίδραση πέψης DNA από κλάσματα του καθαρισμού.

Προσδιορισμός της ενεργότητας κλασμάτωσης της πρωτεΐνης στα στάδια του καθαρισμού έγινε σε ιδανικές συνθήκες για την λειτουργία του ενζύμου {Ausubel et al., 1987}. Ο έλεγχος της ενεργότητας τυποποιήθηκε προκείμενου να είναι συγκρίσιμα τα αποτελέσματα διαφορετικών καθαρισμών. Για κάθε κλάσμα 7μl διαλύματος MDW αναμιγνύονται με 3 μl δείγματος ύστερα από αραίωση (30x / 50x) για την ελαχιστοποίηση της αλλίωσης των συνθηκών αντίδρασης από άλατα έκλουσης και ρυθμιστικές του pH ενώσεις που περιέχονται στα κλασμάτα (n unit assays). Ακολουθεί επώαση στους 37° C για μία ώρα και ανάλυση των προϊόντων της αντίδρασης σε πήκτωμα 1% αγαρόζης (βλ. κεφάλαιο Δ 2.2.). Σαν υπόστρωμα στον προσδιορισμό της ενεργότητας χρησιμοποιήθηκε σε όλες τις περιπτώσεις DNA φάγου λ.

Παρασκευή ρυθμιστικού διαλύματος medium 10x.

medium 10x : 500mM NaCl,
100 mM Tris.Cl (pH 7.5),
100 mM $MgCl_2$,
10 mM DTT.

Διάλυμα MDW v * :

1μl 10x Medium Buffer,
1μgr λ DNA,
συμπλήρωμα σε όγκο 7μl με nanopure νερό.

Δ 3.6. Χρώση πρωτεϊνών με νιτρικό άργυρο, ύστερα από ανάλυση τους σε πηκτώματα ακρυλαμίδης.

Η χρώση πρωτεϊνικών ζωνών σε πηκτώματα ακρυλαμίδης πραγματοποιήθηκε κυρίως με τροποποίηση της μεθόδου που βασίζεται στη χρήση αμμωνιακού διαλύματος νιτρικού αργύρου {Oakley et al., 1980}, {Heukeshoven & Dernick, 1985}. Αποδιατακτικά πηκτώματα ακρυλαμίδης στα οποία προηγήθηκε ηλεκτροφόρηση πρωτεϊνικών δειγμάτων παραμένουν σε 50% μεθανόλη για 2-12 ώρες (fixation). Μετά την απομάκρυνση της μεθανόλης εμβαπτίζονται σε διάλυμα που περιέχει: 60 ml μεθανόλη, 25 ml nanopure νερό, 1.75 ml NH_3 25%, 25 ml NaOH 100mM και 1gr $AgNO_3$. Αν το διάλυμα που προκύπτει δεν είναι διαυγές προσθέτουμε σταγόνες NH_3 μέχρι να το επιτύχουμε. Ο συνολικός όγκος του

διαλύματος είναι 125ml και επαρκεί για ένα πήκτωμα. Το πήκτωμα παραμένει εμβαπτισμένο στο διάλυμα για 1.5 ώρα στο σκοτάδι. Ακολουθεί πλύση με nanopure H_2O με 3-5 αλλαγές για συνολικό χρόνο 8΄. Στην συνέχεια προστίθεται διάλυμα που περιέχει 50mg κιτρικό οξύ και 0.5ml φορμαλδεύδη σε συνολικό όγκο 1lt nanopure H_2O για 5-20΄ με παρακολούθηση της εμφάνισης του πηκτώματος. Τέλος για το σταμάτημα της αντίδρασης απομακρύνεται το προηγούμενο διάλυμα και βάζουμε διάλυμα 50% μεθανόλη και 12% Thiosulfate για συνολικό χρόνο 30΄. Το πήκτωμα στη συνέχεια μπορεί να διατηρηθεί στο σκοτάδι στους 4° C σε νερό. Καθοριστικής σημασίας για την επιτυχία της χρώσης είναι η απουσία αλάτων από το H_2O. Η ευαισθησία της μεθόδου φτάνει στα όρια 1-20ng πρωτεΐνης.

Δ 3.7. Ποσοτικός προσδιορισμός πρωτεΐνών.

Για τον ποσοτικό προσδιορισμό πρωτεϊνικών δειγμάτων από στάδια του καθαρισμού χρησιμοποιήθηκε η μέθοδος Bradford βασισμένη στην πρόσδεση της χρωστικής Coomassie Brilliant Blue G250 {Bradford, 1976}. Για τον προσδιορισμό φτιάχτηκε καμπύλη αναφοράς με πρότυπο διάλυμα 1mg/ml BSA (Bovine Serum Albumin). Η μέθοδος αν και απλή στην χρήση της δεν είναι ικανοποιητικά αξιόπιστη για δείγματα καθαρής πρωτεΐνης.

Δ 3.8. Αντίδραση υπερπέψης DNA (Overdigest).

Σαν μονάδα ενεργότητας ενός περιοριστικού ενζύμου ορίζεται η ποσότητα του ενζύμου που είναι ικανό κάτω από ιδανικές συνθήκες αντίδρασης (Δ 3.5) να υδρολύσει πλήρως 1μg DNA σε μία ώρα στους 37° C. Εκθεση του υποστρώματος για περισσότερο χρόνο ή ποσότητα του ενζύμου μεγαλύτερη απο την απαιτούμενη χαρακτιρίζεται ως υπερπέψη. Μία αντίδραση υπερπέψης χαρακτηρίζεται από τον βαθμό περίσσειας του ενζύμου σε σχέση με την ποσότητα του υποστρώματος και σε σχέση με τον χρόνο αντίδρασης. Η περίσσεια του ενζύμου μπορεί να περιγραφεί ώς : (χρόνος x μονάδες ενζύμου)/ποσότητα DNA σε μg. Από τα προηγούμενα γίνεται φανερό ότι η αντίδραση υπερπέψης μπορεί να επιτευχθεί είτε με την περίσσεια της ποσότητας του ενζύμου είτε με αυξημένο χρόνο αντίδρασης είτε τέλος και με τα δύο. Ετσι υπερπέψη κατά 100 φορές (fold) μπορεί να σημαίνει αντίδραση ενός μg DNA για δώδεκα ώρες παρουσία 8.3 μονάδων του ενζύμου ή γιά 24 ώρες με 4.1 μονάδες του ενζύμου. Συνήθως μια αντίδραση υπερπέψης δεν γίνεται για χρόνο μεγαλύτερο από 12 ώρες λόγω της απώλειας ενεργότητας του ενζύμου στους 37° C.

Δ 4. Κρυστάλλωση και χειρισμός των κρυστάλλων της R.*Pvu*II.

Δ 4.1. Κρυστάλλωση της R.*Pvu*II με την "hanging drop" μέθοδο.

Η "hanging drop" τεχνική είναι μία από τις ευρέως χρησιμοποιούμενες τεχνικές βασισμένες στην αρχή της διάχυσης ατμών (vapor diffusion) {McPherson, 1990}. Μία σταγόνα πρωτεΐνης σε ρυθμιστικό διάλυμα συγκέντρωσης α ενός παράγοντα καθίζησης εξισορροπείται σε κλειστό σύστημα με δεξαμενή όπου η συγκέντρωση του παράγοντα καθίζησης είναι β (β>α). Η σταδιακή μεταβολή της συγκέντρωσης του παράγοντα καθίζησης οδηγεί το πρωτεϊνικό διάλυμα σε κατάσταση υπερκορεσμού με αποτέλεσμα την ελεγχόμενη εναπόθεση πρωτεϊνικού υλικού σε πυρήνες κρυστάλλωσης. Τυπική διαδικασία κρυστάλλωσης της R.*Pvu*II ενδονουκλεάσης με την "hanging drop" μέθοδο είναι η ακόλουθη :

1) 5μl πρωτεΐνης συγκέντρωσης 5-20mg/ml, σε ρυθμιστικό διάλυμα αποθήκευσης (20mM Tris-Cl pH 7.0, 20mM KCl), τοποθετούνται σε σωλήνα eppendorf.

2) Προσθήκη 2 μl ρυθμιστικού διαλύματος οξικού αμμωνίου 500 mM στο επιθυμητό pH (4.8- 5.4), 1 μl θειικού αμμωνίου 3.9 M και 2 μl H_2O. Ακολουθεί ανάμιξη.

3) Σε τρυβλίο πετρί πολλαπλών θέσεων "Linbro" (χωρητικότητας 3.5 ml για κάθε θέση) προσθέτουμε για μία κρυστάλλωση :

- 440 μl θειικό αμμώνιο 100% (3.9M)
- 200 μl Ρυθμιστικό διάλυμα οξικού αμμωνίου 500 mM σε pH ίδιο με της σταγόνας.
- Συμπληρώνουμε με nanopure H_20 μέχρι τα 1000μl.

4) 10 μl από το πρωτεϊνικό μίγμα μεταφέρονται στο μέσο πλαστικής καλυπτρίδας.

5) Το πηγάδι με την δεξαμενή καλύπτεται με την καλυπτρίδα έτσι ώστε η σταγόνα να αιωρείται πάνω από την δεξαμενή. Στεγανοποίηση του πηγαδιού επιτυγχάνεται με την προηγούμενη επάλειψη με λιπαντικό παράγοντα των χειλών του. Η όλη διαδικασία πραγματοποιείται στους 4° C (σε πάγο).

6) Το πιάτο με τις κρυσταλλώσεις μεταφέρεται σε θάλαμο σταθερής θερμοκρασίας 16° C.

Δ 4.2. Μάκρο-σπορά στους κρυστάλλους της R.*Pvu*II. (Macroseeding)

Με την χρήση της "hanging drop" μεθόδου το μέγεθος των κρυστάλλων που παράγονται χαρακτηρίζεται από όρια που καθορίζονται από την ανάπτυξη ατελιών στην επιφάνεια του κρυστάλλου. Με την διαδικασία της μάκρο-σποράς κρύσταλλοι οι οποίοι έχουν παραχθεί με τη "hanging drop" μέθοδο και έχουν

ολοκλήρωση την ανάπτυξη τους μπορούν να συνεχίσουν να μεγαλώνουν. Η μέθοδος προσαρμοσμένη για τους κρυστάλλους της R.*Pvu*II έχει ως εξής :
(Ολη η διαδικασία γίνεται κάτω από μικροσκόπιο.)
- Φρέσκοι κρύσταλοι με ή χωρίς περιορισμένο αριθμό ατελειών μεταφέρονται από το μητρικό διάλυμα σε 100 μl H_20 για χρονικό διάστημα 5-10 λεπτών.
- Ακολουθεί μεταφορά των κρυστάλλων σε νέα ποσότητα 100μl H_2O και αυτό επαναλαμβάνεται 4-5 φορές σε συνολικό χρονικό διάστημα 30΄ (μέχρι να παρατηρηθούν οι πρώτες ενδείξεις διάλυσης της επιφάνειας των κρυστάλλων).
- Οι κρύσταλοι μεταφέρονται σε 100μl διαλύματος σταθεροποίησης (44% θειικού αμμωνίου, 200 mM Tris-Cl pH 5.2). Πραγματοποιούνται 3-4 αλλαγές με διάλυμα ίδιας σύστασης.
- 5μl διαλύματος σταθεροποίησης μαζί με τον κρύσταλλο του οποίου επιθυμούμε την ανάπτυξη μεταφέρονται σε καλυπτρίδα. Στην σταγόνα προστίθενται 5μl φρέσκου πρωτεϊνικού διαλύματος.
- Η καλυπτρίδα μεταφέρεται σε τρυβλίο πετρί όπου έχει προετοιμαστεί δεξαμενή (όγκου 1ml, 22% θειικό αμμώνιο, 100 mM Tris-Cl pH 5.2) έτσι ώστε η σταγόνα να αιωρείται. (βλ. "hanging drops").
- Το τρυβλίο διατηρείται σε σταθερή θερμοκρασία 16° C.

Δ 4.3. Προετοιμασία παραγώγων βαρειών ατόμων.

Για την παραγώγηση των κρυστάλλων της R.*Pvu*II παρασκευάστηκαν διαλύματα ενώσεων των βαρειών ατόμων σε συγκεντρώσεις μεταξύ 1-200mM. Μεταφορά 1μl από τα διαλύματα αυτά σε σταγόνα (συνήθως 10μl) που περιείχε τους προς παραγώγηση κρυστάλλους της R.*Pvu*II έδινε τελική συγκέντρωση μεταξύ 0.1-20mM. Σε περιπτώσεις όπου η παρουσία θειικού αμμωνίου οδηγούσε σε ιζηματοποίηση της ένωσης του βαρέως ατόμου πραγματοποιήθηκε μεταφορά των κρυστάλλων σε σταγόνα που περιείχε 10% PEG8000 σε pH5.0. Οι κρύσταλλοι ύστερα από την χρονικά επιθυμητή παραμονή τους στο διάλυμα με το βαρύ άτομο μεταφερόταν σε τριχοειδή σωληνάρια (capillaries) ύστερα από απομάκρυνση της περίσσειας του μητρικού υγρού.

Δ 4.4. Προετοιμασία κρυστάλλων για έκθεση σε ακτίνες Χ.

Για την πραγματοποιήση της έκθεσης των κρυστάλλων σε ακτίνες-Χ έγινε έγκληση τους σε γυάλινους τριχοειδείς σωλήνες (capillaries) διαμέτρου συνήθως διπλάσιας της δεύτερης σε μέγεθος διάστασης των κρυστάλλων. Απορρόφηση των κρυστάλλων από το μητρικό υγρό στα τριχοειδή σωληνάρια πραγματοποιήθηκε με την συσκεύη Microaspirator Brand κατάλληλη για τον χειρισμό των σωληναρίων.

Ακολούθησε κόψιμο των σωληναρίων ώστε το συνολικό μήκος να μην ξεπερνάει τα 2cm. Η μία άκρη του σωληναρίου κλείστηκε με κερί και η περίσσεια μητρικού υγρού αφαιρέθηκε με σωληνάρια λεπτυμένα με θέρμανση. Οι κρύσταλλοι αφέθηκαν με τόση ποσότητα μητρικού υγρού ώστε μόλις να καλύπεται η επιφάνεια τους. Σε απόσταση 4-5mm αφέθηκαν στις περισσότερες των περιπτώσεων μικρά σταγονίδια μητρικού υγρού, για την αποφυγή της ξύρανσης του κρυστάλλου, και κλείστηκε η δεύτερη άκρη του σωληναρίου. Τέλος τα σωληνάρια με τους κρυστάλλους σταθεροποιήθηκαν πάνω σε γωνιομετρικές κεφαλές με την υποστήριξη πλαστελίνης. Η μορφή των κρυστάλλων (ρσβδοειδείς) ήταν τέτοια ώστε στις περισσότερες περιπτώσεις η μεγάλη διάσταση του κρυστάλλου να είναι κατά μήκος του σωληναρίου.

Δ 5. Προσδιορισμός της δομής της R.PvuII.

Η περιγραφή της μεθοδολογίας που εφαρμόστηκε στα διάφορα στάδια επεξεργασίας των κρυσταλλογραφικών δεδομένων έχει πραγματοποιηθεί στο μεγαλύτερο μέρος της στο κεφάλαιο Β (Αποτελέσματα και συζήτηση). Στή συνέχεια γίνεται συνοπτική αναφορά στα προγράμματα τα οποία χρησιμοποιήθηκαν και λεπτομερειών που κρίνεται απαραίτητο να αναφερθούν και δεν παρουσιάζονται στα αντίστοιχα κεφάλαια των αποτελεσμάτων.

Δ 5.1. Επεξεργασία δεδομένων από δίσκο ειδώλου.

Η επεξεργασία των δεδομένων της φυσικής πρωτεΐνης έγινε με την διαδικασία που περιγράφεται στην συνέχεια και υπήρξε ίδια για όλα τα σύνολα δεδομένων μετρημένα στον X11 σταθμό ακτινοβολίας σύγχροτρον. Ενα είδωλο περιστροφής (oscillation image) από το σύνολο των δεδομένων αναλύθηκε με το πρόγραμμα Imstills και προσδιορίστηκαν οι x,y, z συντεταγμένες των θέσεων των κηλίδων και οι εντάσεις τους. Στην συνέχεια με το πρόγραμμα αυτόματης δεικτοδότησης Refix {Kabsch, 1988} έγινε βελτιστοποίηση των γνωστών ήδη παραμέτρων της στοιχειώδους κυψελίδας με βάση της κηλίδες του ειδώλου που αναλύθηκε και προσδιορίστηκε ο πίνακας κατεύθυνσης (orientation matrix). Χρησιμοποιώντας τις βελτιστοποιημένες παραμέτρους της κυψελίδας και τις παραμέτρους της μέτρησης (γωνία εκκίνησης, βήμα γωνίας, τύπος δίσκου ειδώλου, πίνακας κατεύθυνσης) δημιουργίθηκε με το πρόγραμμα Oscgen αρχείο με όλες τις αναμενόμενες θέσεις κηλίδων για κάθε είδωλο της συλλογής δεδομένων. Στο αρχείο αυτό ενσωματώθηκαν οι εντάσεις των κηλίδων ύστερα από ολοκλήρωση και προσαρμογή με βάση την κατατομή (profile fitting) με το πρόγραμμα Mosflm. Το αρχείο με το σύνολο των ανακλάσεων της μέτρησης χρησιμοποιήθηκε για περισσότερο ακριβή προσδιορισμό του πίνακα κατεύθυνσης με το πρόγραμμα Postcheck με βάση ανακλάσεις μετρημένες σε δύο ή περισσότερα διαδοχικά είδωλα και με τις νέες τιμές για τις παραμέτρους της κυψελίδας και του πίνακα κατεύθυνσης επαναλήφθηκε η διαδιακασία της ολοκλήρωσης. Με το πρόγραμμα Abscale πείραμε αρχείο της μορφής MTZ. Τα προγράμματα Rotavata και Agrovata χρησιμοποιήθηκαν στην συνέχεια για τον υπολογισμό και την εφαρμογή των σχετικών παραγόντων αναγωγής μεταξύ διαφορετικών επικαλυπτόμενων παρτίδων ανακλάσεων {Fox & Holmes, 1965} καθώς και για τον έλεγχο των ισοδύναμων ανακλάσεων και τον υπολογισμό των στατιστικών του συνόλου δεδομένων. Απόριψη ανακλάσεων έγινε με βάση την απόκλιση απο την μέση τιμή των εντάσεων των ισοδύναμων ανακλάσεων. Τέλος έγινε μετατροπή

των δεδομένων από εντάσεις I σε πλάτη παραγόντων δομής F (F^2=I) με το πρόγραμμα Truncate το οποίο παράλληλα τροποποιεί τις αρνητικές τίμες εντάσεων με βάση την στατιστική κατανόμη των εντάσεων του συνόλου των ανακλάσεων {French & Wilson, 1978}. Ολα τα προγράμματα που αναφέρθηκαν ανήκουν στο πακέτο MOSCO/CCP4.

Δ 5.2. Συνθέσεις Patterson και Fourier.

Ολες οι συνθέσεις Patterson και Fourier πραγματοποιήθηκαν με το πρόγραμμα FFT εφαρμόζοντας στον υπολογισμό την προσέγκιση της Fast Fourier Transform {Ten Eyck, 1973}, {Ten Eyck, 1985}. Στίς συνθέσεις Patterson χρησιμοποιήθηκαν ανακλάσεις με $F>2\sigma F$ ή $3\sigma F$ ανάλογα με την ισχύ του συνόλου δεδομένων. Ισόμορφες διαφορές ανακλάσεων μεγαλύτερες από $4*RMS_{iso}$ αποκλείστηκαν από τις συνθέσεις. Οι συνθέσεις πραγματοποιήθηκαν για την χωροομάδα συμμετρίας P_{mmm} (No. 47 στους πίνακες κρυσταλλογραφίας). Οι πυκνότητα των χαρτών ανάχθηκε σε κλίμακα 0-100 ώστε να μπορεί να γίνει άμεση σύγκριση των στατιστικών διαφορετικών χαρτών. Οι συνθέσεις Fourier για τον υπολογισμό των χαρτών ηλεκτρονιακής πυκνότητας με τις MIR φάσεις έγιναν με την χρήση του ίδιου προγράμματος με συμμετοχή στον υπολογισμό όλων των ανακλάσεων μέχρι τα 2.8Å διακριτικότητα.

Δ 5.3. Εύρεση θέσεων των βαρειών ατόμων από συνθέσεις Patterson.

Ο προσδιορισμός των θέσεων των βαρειών ατόμων για τα παράγωγα της πρωτεΐνης έγινε ύστερα από υπολογισμό των συνθέσεων Patterson των ισόμορφών και των ανώμαλων διαφορών σε διακριτικότητα 5 η 3Å όπως έχει ήδη περιγραφεί. Για την αυτόματη ερμηνεία των χαρτών χρησιμοποιήθηκε το πρόγραμμα HASSP με παράγοντα διάκρισης των κορυφών 0.9 (Discrimination factor). Το πρόγραμμα μας έδωσε τις κύριες θέσεις πρόσδεσης για τα παράγωγα Pr, Yb και Hg ενώ για το παράγωγο το Ag μόνο οι θέσεις 1-2 εντοπίστηκαν ως συνδυασμένη λύση (παρουσία όλων των διανυσμάτων (cross vectors)). Παράλληλα οι κορυφές των χαρτών εντοπίστηκαν με το πρόγραμμα PEAKMAX και ελέγχθηκαν για την ορθότητα των λύσεων που προέκυψαν. Οι θέσεις πρόσδεσης των βαρειών ατόμων ελέγχθηκαν για την σύνδεση τους με άξονα τάξεως-2 συμμετρίας με το πρόγραμμα SUPPOSS {CCP4, 1979} το οποίο μας έδωσε τον πίνακα μετασχηματισμού που συνδέει τα δύο μονομερή της R.*Pvu*II.

Δ 5.4. Βελτιστοποίηση των θέσεων των βαρειών ατόμων.

Οι κύριες θέσεις πρόσδεσης για κάθε μονομερές βελτιστοποιήθηκαν ανεξάρτητα με το πρόγραμμα REFINE2 {CCP4, 1979} βασισμένο σε κλασική μεθοδολογία βελτιστοποίησης των παραμέτρων των βαρειών ατόμων {Blow & Matthews, 1973}, {Dickerson, Weinzierl and Palmner, 1968}. Για την βελτιστοποίηση χρησιμοποιήθηκαν μόνο οι ανακλάσεις κεντρικών ζωνών που στην περίπτωση της $P2_12_12$ χωροομάδας είναι οι hk0, h0k και 0hk. Από την διαδικασία της βελτιστοποίησης απορίφθηκαν ανακλάσεις με $F<2\sigma(F)$ καθώς και οι ανακλάσεις ανώτερη εκτίμιση $F_{HUE} < |\langle F_P\rangle - \langle F_{PH}\rangle|$ ή κατώτερη εκτίμιση $F_{HLE} > \langle F_P\rangle + \langle F_{PH}\rangle$. Η βελτιστοποίηση έγινε με βήματα σταδιακά αυξανόμενης διακριτικότητας μέχρι τα όρια διακριτικότητας του συνόλου δεδομένων κάθε παραγώγου. Καθε κύκλος βελτιστοποίησης περιελάμβανε 12 βήματα με εναλλαγή βελτιστοποίησης του παράγοντα θερμοκρασίας B και της πληρώτητας των θέσεων ενώ οι x,y και z συντεταγμένες βελτιστοποιήθηκαν σε όλους τους κύκλους. Οι βελτιστοποιημένες θέσεις του παραγώγου Pr χρησιμοποιήθηκαν για τον υπολογισμό φάσεων της πρωτεΐνης (SIR) με το πρόγραμμα MLPHARE {Otwinowski, 1991} ύστερα από 6 κύκλους βελτιστοποίησης με την "Maximum Likelihood phase" μέθοδο {Otwinowski, 1991},{Dodson & Vijayan, 1971},. Με τις φάσεις της πρωτεΐνης που προέκυψαν έγιναν συνθέσεις Fourier των διαφορών για τα υπόλοιπα παράγωγα. Σε όλες τις περιπτώσεις οι συνθέσεις έδωσαν θέσεις σοδύναμες με αυτές που είχαν ήδη προκύψει από την ανάλυση των συνθέσεων Patterson. Για κάθε θέση πρόσδεσης στην χωροομάδα $P2_12_12$ ορίζονται 8 ισοδύναμες θέσεις που προκύπτουν από μια θέση x,y,z εφαρμόζοντας όλους τους δυνατούς συνδυασμούς για χ=x ή 0.5-x, y=y ή 0.5-y και z=z ή 0.5-z.

Δ 5.5. Προσδιορισμός των φάσεων της πρωτεΐνης.

Οι "βέλτιστες" και οι "πιό πιθανές" φάσεις της πρωτεΐνης προσδιορίστηκαν με το πρόγραμμα MLPHARE στον τελευταίο κύκλο βελτιστοποίησης των παραμέτρων των βαρειών ατόμων. Για τις φάσεις αυτές ο μέσος δείκτης αξιοπιστίας υπήρξε 0.59 για διακριτικότητα μέχρι 2.8Å.

Δ 5.6. Απεικόνηση χαρτών ηλεκτρονιακής πυκνότητας καθώς και μοντέλων της πρωτεΐνης.

Απεικόνισεις, των χαρτών ηλεκρονιακής πυκνότητας και των μοντέλων του μορίου σε οθόνες γραφικών, πραγματοποιήθηκαν με τα προγράμματα FRODO για τον υπολογιστή PS390 της εταιρείας Evans & Satherland και O σε έκδοση για

τον υπολογιστή γραφικών ESV της ίδιας εταιρείας. Χάρτες δημιουργημένοι με το πρόγραμμα FFT μετατράπηκαν με το πρόγραμμα ROMAPPAGE σε μορφή DSN6 κατάλληλη για χρήση με το FRODO. Τα αρχεία συντεταγμένων των μοντέλων ήταν σε μορφή PDB (τύπος: Brookhaven Protein Data Bank). Για την απεικόνιση μορίων χρησιμοποιήθηκε επίσης το πρόγραμμα Sybyl ειδικό για τις διαδικασίες μοντελοποίησης μικρών και μεγάλων μορίων.

Δ 5.7. Τροποποίηση ηλεκτρονιακής πυκνότητας.

Περιοχές αρνητικής πυκνότητας του MIR χάρτη τέθηκαν σε μηδενική τιμή με το πρόγραμμα Truncmap {CCP4, 1979}. Ο χάρτης που προέκυψε χρησιμοποιήθηκε για την δημιουργία αρχείου με όλες τις πιθανές ανακλάσεις όπου τιμές των εντάσεων και των φάσεων προέκυψαν από την αντίστροφη Fourier με το πρόγραμμα SFC {CCP4, 1979}. Ακολούθησε στάθμιση των παραγόντων δομής με το πρόγραμμα HKLWeight {CCP4, 1979} χρησιμοποιώντας το σχήμα $w_i=1-r_i/R$ (R η ακτίνα της εξομάλυνσης και r_i η απόσταση ενός σημείου με πυκνότητα $\rho_i>0$). Στην περίπτωση της R.*Pvu*II χρησιμοποιήσαμε R=10Å επιλογή που έγινε με βάση την διακριτικότητα του MIR χάρτη. (Μια τιμή 8-10 Å είναι τυπική για ένα χάρτη σε 3Å διακριτικότητα.) Στην συνέχεια ακολούθησε αυτόματος προσδιορισμός της περιοχής της ηλεκτρονιακής πυκνότητας που αντιπροσωπεύει την πρωτεΐνη (envelope) για περιεκτικότητα σε διαλύτη 44% (πρόγραμμα Envelope {CCP4, 1979}). Τελικά η τροποποίηση του MIR χάρτη με βάση τον μοριακό φάκελο που υπολογίστηκε έγινε με το πρόγραμμα Flatmap {CCP4, 1979}. Ο εξομαλυνμένος χάρτης που προέκυψε χρησιμοποιήθηκε για επαναμετασχηματισμό σε εντάσεις και φάσεις των ανακλάσεων οι οποίες στην συνέχεια συνδυάστηκαν με το πρόγραμμα Sigmaa {CCP4, 1979} με τίς MIR φάσεις. Η διαδικασία επαναλήφθηκε συνολικά 8 φορές μέχρι να σταθεροποιηθεί ο μέσος δείκτης αξιοπιστίας των φάσεων και είχε ως αποτέλεσμα φάσεις με δείκτη αξιοπιστίας 80%, μετά τον τέταρτο κύκλο η διαδιακασία προσδιορισμού του φακέλου επαναλήφθηκε για περιεκτικότητα διαλύτη 50%.

Για την εφαρμογή της μοριακής εναρμόνισης προσδιορίστηκε αρχικά, με το πρόγραμμα SUPPOS {CCP4, 1979}, ο πίνακας μετασχηματισμού που συνδέει τα δύο σχετιζόμενα με την μή κρυσταλλογραφική συμμετρία μονομερή της R.*Pvu*II. Στη συνέχεια προσδιορίστηκε με οπτική επίβλεψη σε οθόνη γραφικών η περιοχή της στοιχειώδους κυψελίδας και κατα συνέπεια του χάρτη ηλεκτρονιακής πυκνότητας στην οποία έχει εφαρμογή ο άξονας της μή κρυσταλλογραφικής συμμετρίας. Ο χάρτης ηλεκτρονιακής πυκνότητας επεκτάθηκε, με το πρόγραμμα EXTEND {CCP4,1979}, στην περιοχή η οποία μας ενδιέφερε και μετατράπηκε για τις συμμετρίες της χωροομάδας P_1. Η μετατροπή αυτή έγινε για λόγους

επιτάχυνσης των υπολογισμών του προγράμματος εφαρμογής της μεθόδου. Στην συνέχεια ακολουθήσαμε την διαδιακασία που περιγράφεται στο συνοδευτικό κείμενο του προγράμματος SKEWPLANES {Bricogne in CCP4, 1979}, {Bricogne, 1976}.

Δ 5.8. Κρυσταλλογραφική βελτιστοποίηση του μοντέλου της R.PvuII.

Η διαδικασία βελτιστοποίησης του μοντέλου που προέκυψε από τις MIR φάσεις πραγματοποιήθηκε με το πρόγραμμα XPLOR {Brunger, 1990} με την μέθοδο "simulated annealing" ακολουθώντας το πρωτόκολλο "slow-cooling" {Brunger & Krukowski, 1990}. Τα αρχικά βάρη w_a και w_p της στάθμισης μεταξύ του κρυσταλλογραφικού όρου E_{xray} και του χημικού όρου E_{chem} της συνολικής ενέργειας E_{pot} υπολογίστηκαν εκτελώντας μικρής διάρκειας προσομοίωση μοριακών δυναμικών (molecular dynamics). Κάθε στάδιο βελτιστοποίησης με την SA μέθοδο περιλάμβανε πρίν και μετά κυκλους ελαχιστοποίησης ενέργειας με την μέθοδο συζηγούς διαβάθμισης (conjugate gradient) {Powell, 1977}. Στο πρώτο στάδιο η ελαχιστοποίησης της ενέργειας έγινε χρησιμοποιώντας την συνάρτηση ενέργειας απωστικού δυναμικού (repulsive pottential) {Brunger, 1988} για την αποφυγή, λόγω κακής γεωμετρίας του μοντέλου, της ανάπτυξης υπέρμετρα μεγάλων ταχυτήτων των ατόμων εξαιτίας μεγάλων ενεργειών van der Waals (δυναμικά Lennard-Jones). Στην διάρκεια των SA τα φορτία των πλευρικών αλυσίδων των καταλοίπων Lys, Glu, Asp και Arg τέθηκαν σε μηδενικές τιμές καθώς λόγω απουσίας μορίων διαλύτη στην προσομοίωση ο ενεργειακός όρος των ηλεκτροστατικών αλληλεπιδράσεων μεγιστοποιείται. Ο έλεγχος της θερμοκρασίας στην διάρκεια της προσομοίωσης έγινε με την T-coupling μέθοδο {Brunger, 1988} όπου ο συντελεστής τριβής της κίνησης των ατόμων μεταβάλεται με βάση τον λόγο μεταξύ πραγματικής και επιθυμιτής θερμοκρασίας. Η μείωση της θερμοκρασίας πραγματοποιήθηκε σε βήματα των 25° K κάθε 25fs.

Δ 6. Χειρισμός των μοντέλων δομής.

Δ 6.1. Ανάλυση του μοντέλου της R.*Pvu*II.

Το πρόγραμμα Contact {CCP4, 1979} χρησιμοποιήθηκε για τον υπολογισμό αποστάσεων μεταξύ ατόμων που δυνητικά συμμετέχουν στην δημιουργία υδρογονικών δεσμών ή γεφυρών αλάτων του μοντέλου της R.*Pvu*II. Οι αποστάσεις περιορίστηκαν μεταξύ 2.5-3.5Å και οι γωνίες NH···O>100°. Τα γεωμετρικά κριτήρια των υδρογονικών δεσμών βασίστηκαν σε στατιστικές απο επιλυμένες πρωτεϊνικές δομές {Baker & Habbard, 1986}. Για τον εντοπισμό των

υδρόφοβων αλληλεπιδράσεων μεταξύ των δύο μονομερών επιλέχθηκαν αποστάσεις ατόμου-ατόμου μικρότερες των 5Å με το πρόγραμμα Contact {BIOMOL}. Σαν υδρόφοβες αλληλεπιδράσεις θεωρήθηκαν επαφές C-C καταλοίπων με αλειφατική πλευρική αλυσίδα ή πλευρική αλυσίδα με μήκος που να επιτρέπει σε μέρος της να μετέχει σε υδρόφοβες αλληλεπιδράσεις (Lys). Με το ίδιο πρόγραμμα εντοπίστηκαν τα κατάλοιπα των οποίων άτομα των πλευρικών αλυσίδων πλησιάζουν σε απόσταση μικρότερη των 4Å από θέσεις πρόσδεσης βαρειών ατόμων. Στατιστικές των παραγόντων θερμοκρασιας B υπολογίστηκαν με το πρόγραμμα Baverage {CCP4, 1979}. Ο χαρακτηρισμός περιοχών του μορίου ως α-έλικες και β-κλώνους έγινε με το πρόγραμμα DSSP {CCP4, 1979}.

Δ 6.2. Συγκρίσεις μεταξύ μοντέλων των R.*Pvu*II και R.*Eco*RV, R.*Pvu*II σε σύμπλοκο με DNA.

Ο αρχικός υπολογισμός των πινάκων μετασχηματισμού μεταξύ δομικά ομόλογων μοντέλων πραγματοποιήθηκε το υποπρόγραμμα LSQ_EXPLICIT του O καθορίζοντας συγκεκριμένα ζεύγη C^{α} ατόμων των προς σύγκριση μορίων. Οι πίνακες μετασχηματισμού βελτιώθηκαν με το υποπρόγραμμα LSQ_IMPROVE επίσης του O. Η διαδικασία βελτιστοποίησης περιλάμβάνει 10 κύκλους όπου ο πίνακας μετασχηματισμού βελτιώνεται παράλληλα με την ένταξη στην στοίχηση δομικών περιοχών που στοιχίζονται, με βάση τον εκάστοτε προηγούμενο πίνακα, καλύτερα από ένα όριο rms αποστάσεων. Ο πίνακας που προκύπτει μπορεί να εφαρμοστεί στις συντεταγμένες ενός μορίου με το υποπρόγραμμα LSQ_molecule. Η όλη διαδικασία περιγράφεται στην αναφορά {Jones & Kjeldgaard, 1992}.

Προσάρτημα Ι

Δημοσιεύσεις

 © 1990 Oxford University Press

Complete nucleotide sequence of the *Pvu* II restriction enzyme gene from *Proteus vulgaris*

A.Athanasiadis, M.Gregoriu, D.Thanos, M.Kokkinidis and J.Papamatheakis
University of Crete, Department of Biology and Institute of Molecular Biology and Biotechnology (IMBB), PO Box 1527, GR-71110 Heraklion, Crete, Greece

Submitted September 26, 1990 EMBL accession no X52681

The gene coding for the *Pvu* II restriction enzyme was isolated by screening pools of pBR322 for the presence of enzyme activity. We present its complete nucleotide sequence which contains an open reading frame of 157 amino acids. The approximate molecular weight of 18kDa is in good agreement , SDS-PAGE determined molecular weight of the protein. The location of the open reading frame is consistent with deletion analysis data (1). Computer-predicted promoter sequences are underlined, the Shine-Dalgarno signal is doubly underlined. The presented sequence overlaps partially with the methylase gene recently published (2). The region upstream of the *Pvu* II-R gene contains also the promoter elements for the methylase gene which is transcribed in the opposite direction (1). No homology between the *Pvu* II-R gene sequence and other known restriction enzyme sequences was found.

ACKNOWLEDGEMENTS

We thank Prof. V. Bouriotis for his valuable suggestions and the Enzyme Technology Group of IMBB for their technical assistance.

REFERENCES

1. Tao,T , Walter,J., Brennan,K J., Cotterman,M M and Blumenthal,R M (1989) *Nucl. Acids Res.* 17, 4161–4175
2. Blumenthal,R M , Gregory,S A. and Cooperider,J S (1985) *J of Bact* 164, 501–509.

```
   1 ACTAGTTGTGTAGGCAGGTTTTTTTCCAAAATTCAACATATCATTGCTACTCATAGTCTGTAGATTCAAAGTCATCATAC
  81 CATCATTATCCCGTCTATGAGCAGAAATACAAATCCTTTATCAGCCCGATTAACCCTTGCAAAGAATGTAAAAAAAATGC
 161 GAGGCGAGCTAGGTCTATCCCAAGAAAGCTTAGCTGATCTAGTGGGAATCCATAGAACCTACATTGGTTCAATTGAACGA
 241 GCGGAAAGGAATATATCGATAGACAACATTGAGCGAATAGCAAATGCCTTAAATGTTTCTATATCAATACTAATGATGGA
                M  S  H  P  D  L  N  K  L  L  E  L  W  P  H  I  Q  E  Y  Q  D  L  A  L
 321 ACACGAAAATGAGTCACCCAGATCTAAATAAATTATTAGAGCTTTGGCCGCATATACAGGAATATCAAGACTTAGCATTA
      K  H  G  I  N  D  I  F  Q  D  N  G  G  K  L  L  Q  V  L  L  I  T  G  L  T  V  L
 401 AAACATGGAATAAATGATATTTTTCAAGATAATGGTGGAAAGTTGCTTCAAGTCCTTCTAATTACAGGATTAACAGTACT
       P  G  A  E  G  N  D  A  V  D  N  A  G  Q  E  Y  E  L  K  S  I  N  I  D  L  T
 481 ACCAGGACGAGAAGGTAATGATGCTGTAGATAACGCAGGACAAGAATACGAGTTAAAATCAATAAACATAGACCTCACTA
     K  G  F  S  T  H  H  H  M  N  P  V  I  I  A  K  Y  A  Q  V  P  W  I  F  A  I  Y
 561 AAGGTTTTTCAACTCACCACCACATGAATCCTGTAATTATTGCAAAATATAGACAAGTACCTTGGATTTTTGCCATATAC
      A  G  I  A  I  E  A  I  Y  A  L  E  P  K  D  L  E  F  Y  Y  D  K  W  E  A  K  W
 641 CGTGGTATCGCAATAGAAGCTATATACAGATTAGAGCCAAAAGATCTAGAATTTTACTATGATAAATGGGAAAGGAAATG
       Y  S  D  G  H  K  D  I  N  N  P  K  I  P  V  K  Y  V  M  E  H  G  T  K  I  Y
 721 GTATTCAGATGGGCATAAAGATATTAACAACCCTAAAATACCTGTAAAATATGTAATGGAACATGGGACAAAGATTTACT
     *
 801 AAATTGGAGCTACATTCATGGTTCCGATAAGACCCAATTATGTTAAACGGGCGAGTTTACACCTTAAAACCGCCCGTCAG
 881 TCACCATCAGAACGCATCAGCACGATTTTAAGCACCAAACACCCCCCATAACACCCAAATCCATCCTGAAAGTTTATAAC
 961 GGTTTCTGTGACGTTTGAGGGCTGTTTACATCCGTTTTTCGTCCGACTTTGATCGCCTAAACCGATGAAAGTCGCAAACT
1041 TGTTTGTGGCTTGAAGCGGTTTTATCCTGTTAAAAACCGATACTGAAACCCCCGACTACGACTCTGCTCCTGTCGCTGGA
1121 TCTGCTGTGTCAGCCGTTCTGATACCTGTTCAGCAGTCCGGTTAAAGTCGCTACTGTTGACTGATTTCCCTGTATTGCAT
1201 GGCTAACTGACTGTTCTGTGTCTCGAGTTCCCCCGACACGCTGAGATAAATTGCTGCATAACGGTCTCTTGCTGTTGTAA
1281 CAAGGTGTTCAAGGCGTTCTCTAAGAGTTGTTCTGTCTCGGTCA
```

Reprinted from *J. Mol Biol* (1991) **222**, 451–453

Purification, Crystallization and Preliminary X-ray Diffraction Studies of the *Pvu*II Endonuclease

Alekos Athanasiadis and Michael Kokkinidis

J. Mol Biol (1991) 222, 451–453

Purification, Crystallization and Preliminary X-ray Diffraction Studies of the *Pvu*II Endonuclease

Alekos Athanasiadis and Michael Kokkinidis

University of Crete, Department of Biology
and
Institute of Molecular Biology and Biotechnology (IMBB)
P.O. Box 1527, GR-71110, Heraklion, Crete, Greece

(Received 1 July 1991; accepted 19 August 1991)

The *Pvu*II endonuclease (*Pvu*IIR) is a restriction enzyme from a type II restriction-modification system of *Proteus vulgaris* coded on plasmid pPvu1. The protein recognizes the DNA sequence 5′ CAG′CTG 3′ and shows no sequence homology to other restriction enzymes. This makes *Pvu*IIR an interesting subject for structural determination. A purification procedure was developed that yields milligram quantities of the *Pvu*IIR from plasmids expressed in the *Escherichia coli* strain HB101. The protein was crystallized using ammonium sulphate as precipitant. The crystals are orthorhombic, space group $P2_12_12$ with cell dimensions: $a = 84{\cdot}2$ Å $b = 106{\cdot}2$ Å, $c = 46{\cdot}9$ Å. The asymmetric unit contains one *Pvu*IIR dimer Diffraction extends to 2·3 Å, so the crystals may permit structural determination at atomic resolution.

Keywords restriction-modification systems, *Pvu*II endonuclease; enzyme purification, protein crystallization, hanging-drop technique

Type II restriction-modification systems (RMS2s†), of which more than 1300 have been identified, provide attractive model systems for the analysis of the structural basis of the sequence-specific recognition of double-stranded DNA by proteins. A typical RMS2 consists of two enzymatic activities, a methyltransferase and an endonuclease activity (Wilson, 1988) The two activities are associated with separate proteins, both enzymes recognize the same DNA sequence, which may vary in size from four to ten nucleotides The major function of RMS2s is assumed to be defence against bacteriophage DNA (Kruger & Bickle, 1983). They have gained major importance in molecular biology because they have been used in cloning and other techniques

*Pvu*II, a type II restriction-modification system of the *Proteus vulgaris* Gram-negative bacteria consists of two plasmid-encoded genes (Blumenthal *et al.*, 1985). The two genes lie adjacent and are transcribed divergently, leaving a small region of DNA between them that contains the promoter elements of the genes (Blumenthal *et al.*, 1985). There is recent evidence that an additional, small open reading frame (ORF) in the intergenic region plays a role by coding a *trans*-acting regulatory protein for the *Pvu*II endonuclease gene (Tao *et al.*, 1991).

The *Pvu*II endonuclease (*Pvu*IIR) is a protein of 157 residues with a predicted M_r of 18,345; its complete sequence has been recently reported and shows no homology to other known proteins (Athanasiadis *et al.* 1990) Gel filtration studies have shown that *Pvu*IIR exists in solution in a dimeric form (unpublished results) This is consistent with results obtained with other restriction enzymes, which indicate that the homodimer is the catalytically active form of the protein (Jack *et al.*, 1991; Wilson, 1988). Catalytically active *Pvu*IIR requires Mg^{2+} (Wilson, 1988) and recognizes the palindromic sequence 5′ CAG′CTG 3′, which is cleaved in the middle, leaving blunt-ended DNA fragments (Gingeras *et al.*, 1981) Star activity may appear at high enzyme concentrations (unpublished results)

To date, despite the large number of known restriction enzymes, only the crystal structure of *Eco*RI (Kim *et al.* 1990) is available, crystallographic work on *Bam*HI (Jack *et al.*, 1991) and *Eco*RV (D'Arcy *et al.*, 1985) is in progress. The catalytic activity of *Pvu*IIR, its lack of homology to other proteins and the very limited amount of structural information about restriction enzymes makes

† Abbreviations used: RMS, restriction-modification system; *Pvu*IIR, *Pvu*II endonuclease

0022-2836/91/230451-03 $03·00/0

© 1991 Academic Press Limited

the crystallographic analysis of the protein an interesting subject

In order to obtain sufficient amounts of *Pvu*IIR at levels of purity that are suitable for crystallization, the following scheme for production and purification of the protein was developed *Escherichia coli* cells (strain HB101) harboring the plasmid pPvuI region, which contains both genes of the *Pvu*II restriction-modification system, in the pPUC18 vector were used for the production of the protein. The bacteria were cultured in a 30 liter fermenter using standard nutrients and conditions for *E. coli* growth. Approximately 300 g of cell paste was harvested in each fermentation and stored at −70°C For a typical preparation 80 to 100 g of frozen cell paste was thawed and suspended in 50 mM-Tris · HCl (pH 7·2) 1 mM-EDTA and 10 mM-2-mercaptoethanol. Crude extracts were prepared in the presence of the protease inhibitors benzamidine, aprotinin, phenylmethylsulfonyl fluoride and leupeptin using sonication or a French press After removal of nucleic acids with 0·5% polyethylenimine (PEI) and fractionation with ammonium sulfate (saturation range 30% to 70%), the extracts were dialyzed against buffer A (20 mM-potassium phosphate (pH 7·4), 1 mM-EDTA, 100 mM-KCl, 10 mM-2-mercaptoethanol). The extracts were loaded on a 300 ml phosphocellulose column equilibrated with buffer A. After thorough washing with the same buffer, the column was eluted with a two liter linear gradient of KCl (0·1 M to 0·6 M) in buffer A. *Pvu*IIR eluted at approximately 300 mM-KCl. The catalytically active fractions were further purified by fast liquid protein chromatography, anion-exchange chromatography using an 8 ml MonoQ column equilibrated with buffer B (20 mM-Tris · HCl (pH 7·4), 1 mM-EDTA, 10 mM-2-mercaptoethanol) containing 10 mM-KCl A 160 ml linear gradient (10 mM to 600 mM) of KCl in buffer B was applied; *Pvu*IIR was eluted at about 100 mM-KCl Pure *Pvu*IIR (at least 95% of the protein present, as judged from a silver-stained SDS/polyacrylamide gel) was obtained by gel filtration on a 500 ml Sephadex G-75 column equilibrated with buffer B containing 100 mM-KCl The final protein preparation was dialyzed against 10 mM-Tris · HCl (pH 7·4) containing 5 mM-KCl and concentrated to approximately 5 mg/ml using Amicon Centricon 10 microconcentrators. The protein concentration was determined using the Bradford (1976) assay with bovine serum albumin as a standard. A total of 5 mg of pure protein was the final yield from 80 to 100 g of cell paste The MonoQ run was performed at room temperature, all other purification steps were performed at 4°C

Crystallization trials were performed using the hanging-drop vapor diffusion method. The best results were obtained using ammonium sulfate as precipitant. The 10 μl droplets contained a 2·5 mg/ml protein solution buffered at pH 5·0 with 100 mM-ammonium acetate and 10% to 20% ammonium sulfate. The droplets were equilibrated at 18°C against 1 ml of reservoir solution containing 20% to 50% ammonium sulfate The best results were reproducibly obtained at a concentration of 44% ammonium sulfate. Well-formed, parallelepiped-shaped crystals, started to appear two days after preparation of the droplets, in the presence of protein precipitate and phase separation, and reached a size of approximately 0·5 mm in their longest dimension after almost two weeks

Crystals with the approximate dimensions of 0·5 mm × 0·3 mm × 0·2 mm were sealed in glass capillaries and exposed to X-rays using an Enraf-Nonius CAD4 diffractometer operating at 40 kV and 32 mA (room temperature, Ni-filtered CuK_α radiation). Unit cell parameters were determined from 25 reflections recorded in the resolution range between 4 Å and 3 Å. The crystals are orthorhombic with unit cell dimensions $a = 84{\cdot}2$ Å, $b = 106{\cdot}2$ Å and $c = 46{\cdot}9$ Å The conditions for systematic absences $h00$ $h \neq 2n$ and $0k0$ $k \neq 2n$, identified the space group as $P2_12_12$. The crystals diffract to 2 3 Å and are not particularly sensitive to X-ray irradiation. Calculation of V_m as defined by Mathews (1968) gives a value of approximately 2·9 $Å^3$/dalton, if the M_r of the asymmetric unit is assumed to correspond to one *Pvu*IIR dimer This value of V_m is within the experimentally observed range (Mathews, 1968). The presence of one dimer in the asymmetric unit is consistent with gel-filtration experiments, which indicate that *Pvu*IIR behaves as a dimer in solution. To date, a complete set of diffraction data to a resolution of 4·0 Å has been collected from a native crystal on the diffractometer mentioned above, 88% of the collected reflections have $I > 3\sigma(I)$, where I is the X-ray reflection intensity. We are in the process of collecting native data to a higher resolution and we have initiated a search for suitable heavy-atom derivatives that will permit structure determination by the multiple isomorphous replacement method.

We thank Dr V Bouriotis, Mrs M Pagomenou and the staff of the Enzyme Technology Group of IMBB for helpful assistance

References

Athanasiadis, A , Gregoriu, M , Thanos, D , Kokkinidis, M & Papamatheakis, J. (1990) Complete nucleotide sequence of the *Pvu*II restriction enzyme gene from *Proteus vulgaris* *Nucl. Acids Res* **18** 6434

Blumenthal, R. M , Gregory, S A & Cooperider, J S (1985) Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analysing a methylase-sensitive phenotype in *Escherichia coli* *J Bacteriol* **164** 501–509

Bradford M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of dye binding *Anal Biochem* **72**, 248–254

D'Arcy, A., Brown, R S., Zabeau M , van Resand R W & Winkler, F K. (1985). Purification and crystallization of the *Eco*RV restriction endonuclease *J Biol Chem.* **260**, 1987–1990

Gingeras, T R., Greenough, L., Schildkraut L & Roberts, R. I (1981) Two new restriction endo-

nucleases from *Proteus vulgaris* *Nucl Acids Res* **9**, 4525–4536

Jack, W E, Greenough, L, Dourner, L, Xu, S-Y, Strzelecka, T, Aggarwal, A K & Schildkraut, I (1991) Overexpression, purification and crystallization of *Bam*HI endonuclease *Nucl Acids Res* **19**, 1825–1829

Kim, Y C, Grable, J C, Love, R Greene, P J & Rosenberg, J M (1990) Refinement of the *Eco*RI endonuclease crystal structure: A revised protein chain tracing *Science*, **249**, 1307–1309

Kruger, D H & Bickle, T A (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev* **47**, 345–360

Mathews B W (1968). Solvent content of protein crystals *J. Mol. Biol.* **33**, 491–497

Tao, T Boume, J C & Blumenthal R M (1991) A family of regulatory genes associated with type II restriction-modification systems *J. Bacteriol* **173** 1367–1375

Wilson G G (1988). Type II restriction-modification systems. *Trends Genet* **4**, 314–318

Edited by R Huber

Crystal structure of *Pvu*II endonuclease reveals extensive structural homologies to *Eco*RV

A. Athanasiadis, M. Vlassi, D. Kotsifaki, P. A. Tucker[1], K. S. Wilson[2] and M. Kokkinidis[3]

The crystal structure of the dimeric *Pvu*II restriction endonuclease (R.*Pvu*II) has been determined at a resolution of 2.4Å. The protein has a mixed α/β architecture and consists of two subdomains. Despite a lack of sequence homology, extensive structural similarities exist between one R.*Pvu*II subdomain and the DNA-binding subdomain of *Eco*RV endonuclease (R.*Eco*RV); the dimerization subdomains are unrelated. Whithin the similar domains, flexible segments of R.*Pvu*II are topologically equivalent to the DNA-binding turns of R.*Eco*RV; potential catalytic residues can be deduced from the structural similarities to R.*Eco*RV. Conformational flexibility is important for the interaction with DNA. A possible classification of endonuclease structures on the basis of the positions of the scissile phosphates is discussed.

Department of Biology, University of Crete and Institute of Molecular Biology and Biotechnology (IMBB), University of Crete, PO Box 1527, GR-71110 Heraklion, Crete, Greece
[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany
[2]European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany
[3]To whom correspondence should be addressed

Bacterial type II restriction–modification systems (RMS) are attractive model systems for studying the structural basis of protein-DNA interactions. A typical type II RMS consists of two proteins, an endonuclease and a methyltransferase. Each protein of a pair binds specifically to the same DNA sequence which may vary in size from four to eight base pairs. Once bound to a site, endonucleases cleave both DNA strands if Mg^{2+} is present[1,2], whereas methyltransferases transfer a methyl group from the cofactor S-adenosylmethionine to one of the bases of the recognition sequence, thus protecting the bacterium's DNA from its own endonucleases[1,2]. A major function of type II RMSs is assumed to be defence against foreign, unmethylated DNA, such as that of viruses[1–4]. Much less is known about other possible functions which include DNA recombination and repair[1,4]. Of the more than 2,100 type II restriction endonucleases discovered so far[5], very few have been studied in any detail. Although many have the same specificity, around 200 different DNA specificities have been identified[5]. At the level of amino acid sequence, restriction endonucleases are quite a diverse; it is possible, therefore, that the different specificities represent a variety of strategies for recognizing DNA. The structure–function relationships of restriction endonucleases are thus a potentially rich source of information about protein–DNA interactions.

The *Pvu*II endonuclease (R.*Pvu*II) from the *Pvu*II RMS of *Proteus vulgaris* has 157 residues of known sequence[6]; protein produced from the bacterial host lacks the N-terminal methionine[7]. R.*Pvu*II recognizes the palindromic sequence 5′ CAG↓CTG 3′ which is cleaved in the middle, leaving blunt-ended DNA fragments[8]. The active form is the dimer and, like other type II endonucleases, R.*Pvu*II requires Mg^{2+} for DNA cleavage[9,10].

To date, the only type II endonucleases for which three-dimensional structures have been reported are R.*Eco*RI[11,12] and R.*Eco*RV[13,14]. Despite the similarity of their active sites, the two enzymes are structurally dissimilar and use different strategies to attack DNA; these have been related to differences in the positions of the scissile phosphodiester bonds[15]. Several important questions related to DNA recognition and the catalytic mechanism are still open[13]. The poorly understood endonuclease structure–function relationships, the lack of homology to other proteins[6,13] and the major importance of endonucleases in molecular biology in cloning and other techniques, make R.*Pvu*II an interesting subject for crystallographic analysis.

Protein structure

The 3D-model of R.*Pvu*II has been refined to a resolution of 2.4 Å. Electron density maps (Fig. 1) of excellent quality show unambiguously the positions of all residues, with the exception of Arg 54 and Glu 55 which are disordered and have been omitted from the final model. A ribbon representation[16] of the crystal structure is shown in Fig. 2. The connectivity of the secondary structural elements of the protein is shown in Fig. 3*a*. The U-shaped dimer (approximate dimensions 70 x 55 x 30 Å) has a local dyad corresponding to a rotation of 179°. The two monomers of 156 amino acids (numbered 2–157) show

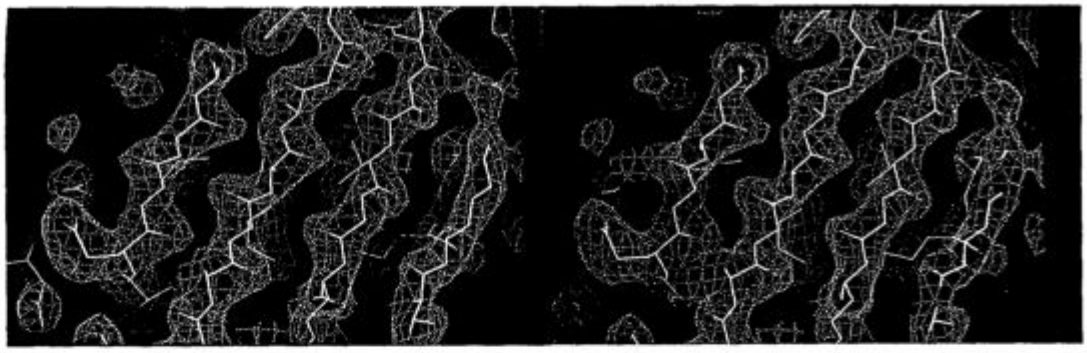


Fig. 1 The final $2F_o$-F_c electron density map of *Pvu*II (contoured at 1σ) showing part of the central β-sheet.

small differences corresponding to an r.m.s.d. of 0.4 Å for the backbone atoms (excluding Ser 2-Pro 4, Gly 53, Gly 56 which exhibit the highest deviations). R.*Pvu*II has a mixed α/β molecular architecture (Fig. 2*a*): the central part consists of a mixed parallel/antiparallel β-sheet (Fig. 1) comprising three long β-strands (βc, βe and βf). The sheet is extended on one side by two antiparallel short strands (βa and βb) which are connected by a disordered loop (residues Arg 54, Glu 55) and on the other side by a further short strand (βh). On the top face of this sheet are two short β-strands (βd and βg) and three α-helices (αC, αD and αE). On the bottom face of the central β-sheet are positioned α-helices *a*A and αB. Helix αA forms the bottom of the molecule and because of a Pro residue in a central position, it is kinked and makes extensive contacts with helices αA and αB of the other monomer, forming a large part of the dimer interface (Fig. 2*b*). The monomer is organized into two independent subdomains. Subdomain A (Ser 2-Asn 35) which consists of helix αA and the loop (His 26-Asn 35) connecting helices αA and *a*B, forms most of the dimer interface and will be referred to as the 'dimerization' subdomain. Subdomain B comprises the rest of the molecule. Dimerization interactions involve subdomain A and parts of helix αB: For each monomer, the carboxy-terminal part of helix αA (residues following the central Pro), helix αB and the N-terminal part of helix αA from the other monomer form a three-helix bundle with extensive hydrophobic interactions in its core. The dimer is furthermore stabilized by hydrogen bonds and salt bridges between helices in the dimerization interface and two hydrogen-bonded loops which are related by the local dyad and connect in each monomer helices αA and αB.

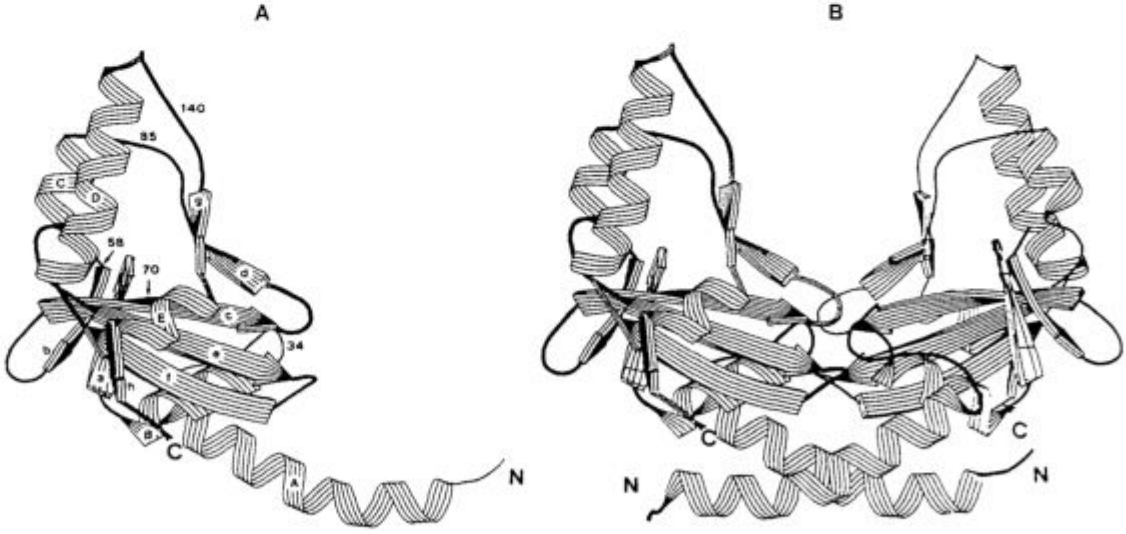


Fig. 2 R *Pvu*II ribbon representation[16] *a*, Monomer structure with α-helices and β-strands labelled N and C indicate the N and C termini of the protein The approximate positions of selected residues are indicated with residue numbers *b*, Dimer structure

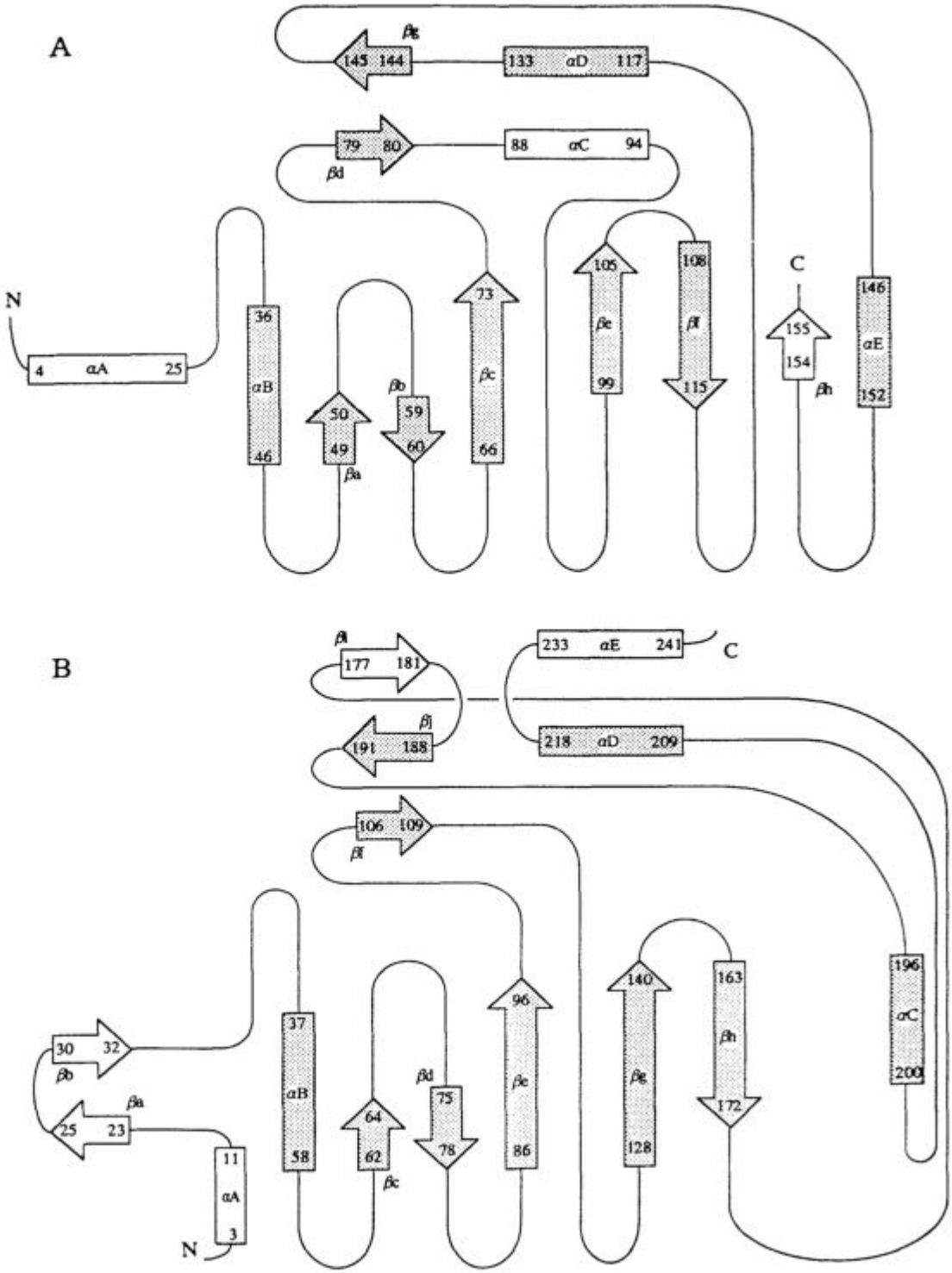


Fig. 3 Topological representation of *a*, R.*Pvu*II and *b*, R.*Eco*RV. Secondary structural elements of each protein that have topologically equivalent counterparts in the other enzyme are indicated by shading. In R.*Pvu*II secondary structure labelling is as for Fig 2, in R.*Eco*RV the labelling is as in ref 14.

Table 1 Data collection, MIRAS phasing and refinement

	Native	Pr[a]	Ag[b]	Hg[c]	Yb[d]	Hg1[b]
Data collection						
Source	X11	GX21	GX21	X11	GX21	GX21
Maximum resolution	2 3	2 5	3.0	2.8	2 5	2 5
Completeness (%)	96.3	78.0	98.6	95 1	83 9	85 8
Unique reflections	16,315	11,985	8,886	11,031	13,804	13,173
No. of reflections with anomalous data	—	9,147	5,658	8,964	6,329	—
R_{sym}[e]	7 6	8.8	7 4	4 0	6.9	9 3
MIRAS analysis						
Mean isomorphous difference[f]	—	24 2	22 5	22.5	30.0	25 6
K_{amp}[g]	—	3.04	4.70	8 92	6 14	_
No of major sites	—	2	5	2	2	2
Phasing power[h]	—	1.85	0 81	1.11	1 70	1 05
R_{Cullis}[i]	—	0 82	0.92	0 91	0.86	0.92
Refinement						
Resolution limits (Å)		6–2 4				
Reflections with F>2σ		15,332				
R factor[j] (%)		18.4				
Protein atoms included		2,544				
Waters added		79				
R.m.s. in bond lengths (Å)		0.023				
R m s. in bond angles (degree)		3.6				

[a]$Pr(NO_3)_3$,
[b]$AgNO_3$,
[c]Thiomersal[19],
[d]$Yb_2(SO_4)_3$
[e]$R_{sym}=\Sigma_h\Sigma_i|I_{h,i}-I_h|/\Sigma_h\Sigma_i I_{h,i}$ where I_h is the mean of the i observations of reflection h
[f]Mean isomorphous difference$=\Sigma|F_{PH}-F_P|/\Sigma F_{PH}$ where F_{PH} and F_P are the derivative and native structure factor amplitudes respectively
[g]$K_{amp}=2(<(F_{PH}-F_P)^2>/<(F_{PH}(+)-F_{PH}(-))^2>)^{1/2}$
[h]Phasing power$=<F_{H(calc)}>/<|F_{PH(obs)}-F_{PH(calc)}|>$
[i]$R_{Cullis}=\Sigma||F_{PH}\pm F_P|-F_{H(calc)}|/\Sigma|F_{PH}-F_P|$ for centric reflections where $F_{H(calc)}$ is the calculated heavy atom structure factor.
[j]R factor$=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$. The r m.s. in bond lengths and bond angles is the r.m.s d. of bond lengths and bond angles from ideal values

Relationship to other endonucleases

Sequence comparisons of R.*Pvu*II to other proteins produce alignments of marginal significance. In particular, no homologies exist to other endonucleases of known 3D-structure. R.*Pvu*II is structurally unrelated to R.*Eco*RI[11,12], with the exception of similarities in the central β-sheet which forms the main part of the core in both structures. Extensive topological and structural homologies exist however between R.*Pvu*II and R.*Eco*RV[14] (Figs 3 and 4): spatial alignment of the two enzymes reveals a set of secondary structure elements that is present in both structures in topologically equivalent positions (Fig. 3). In R.*Pvu*II this set includes α-helices αB,αD,αE and β-strands βa,βb,βc,βd,βe,βf,βg (Figs 3*a* and 4*a*); they are topologically equivalent to α-helices αB,αD,αC and β-strands βc,βd,βe,βf,βg,βh,βj respectively in R.*Eco*RV (Fig. 3*b*). The R.*Pvu*II helix αB is equivalent only to the N-terminal part of the kinked R.*Eco*RV helix αB. Subdomain B is thus extensively related to the DNA-binding subdomain of R.*Eco*RV[14], although there is no relation between the dimerization subdomains of the two proteins. After spatial alignment most of the secondary structure elements that are topologically equivalent superimpose completely or partially : 46 R.*Pvu*II residues (around 30% of the protein) from α-helices αB,αD,αC and β-strands βc,βd,βe,βf,βg superimpose with residues from the equivalent R.*Eco*RV segments with an r.m.s d. in the Cα positions which is better than 2 5 Å (Fig. 4*a*,*b*). The order in which equivalent secondary structure elements occur in the sequences is the same for both proteins with the exception of a segment comprising αD,βg and αE in R.*Pvu*II. The rearrangement in this segment implies that alternative connectivities between equivalent segments are possible in the two structures. Topological equivalences are reflected to some extent in the sequences: α-helices αD in R.*Pvu*II and R.*Eco*RV are equivalent and both have an unusually high content of aromatic residues.

Relationship to R.*Eco*RV

On the basis of spatial alignments, regions in R.*Pvu*II can be identified which correspond to the DNA-binding turns of R.*Eco*RV: the disordered loop between βa and βb is equivalent to the R.*Eco*RV DNA binding 'Q-turn'[14], whereas a region comprising a segment with three successive His residues (His 83,His 84,His 85) and a seg-

ment with two successive Asn residues (Asn 140, Asn 141), corresponds to the DNA recognition 'R-turn'[14] in R.*Eco*RV; if dimers are aligned, His 84 superimposes with R *Eco*RV residues that are involved in base-specific DNA recognition. The three R.*Pvu*II segments that are structurally related to the DNA-binding loops of R.*Eco*RV, are rich in residues which can bind to DNA through salt bridge or hydrogen bond formation[17]. Flexibility, as reflected in temperature factors, is a further common feature of the above R.*Pvu*II segments and the DNA-binding turns in R.*Eco*RV.

The R *Eco*RV catalytic residues Asp 74, Asp 90 and Lys 92 superimpose very well with R.*Pvu*II residues Asp 58, Glu 68 and Lys 70; the segments of the two structures to which these residues belong are topologically equivalent. Furthermore, there is a striking conservation at structural and sequence level between these three residues and the catalytic residues of R.*Eco*RI. This suggests that Asp 58, Glu 68 and Lys 70 are possibly part of the catalytic site in R.*Pvu*II. It was proposed earlier[15] that catalytic residues in type II endonucleases define a sequence motif of the form ...PDX$_n$(D/E)ZK... (n=10–30; X, Z can be any amino acid). If Asp 58 is part of the R.*Pvu*II catalytic site, then the strict requirement of a Pro preceding the first Asp in this motif can be dropped; at the structural level Asp 58 is preceded by a turn. This agrees with our analysis of type II endonuclease sequences (M. Kokkinidis, unpublished material) which shows that for endonucleases that leave blunt ends, the first Asp of the motif is usually preceded by residues (including Pro) with a high potential for turn formation[18]. Our analysis also shows that residue Z is overwhelmingly hydrophobic or has occasionally a long side chain (Gln, Arg). An interpretation of this behaviour is possible if the structural homologies between R.*Pvu*II and R.*Eco*RV extend also to other endonucleases: in R.*Pvu*II and R.*Eco*RV, Z is located in an outer strand of the central β-sheet and faces the hydrophobic core; this position can be best occupied either by a hydrophobic residue or by Gln or Arg because long side chains can escape to the surface of the protein while still exposing a hydrophobic part to the core. In R.*Eco*RV there is an acidic residue (Glu 45) close to the active site which on the basis of mutagenesis studies[14] seems to be important for catalysis. This residue has no counterpart in R.*Eco*RI, whereas its counterpart in the structure of R.*Pvu*II is the hydrophobic residue Leu 39. It may be thus concluded that Glu 45 is not an active site residue, but its precise role in R.*Eco*RV remains to be determined.

DNA binding and free enzyme conformation

The cleft of the U-shaped dimer is considerably wider in R.*Pvu*II than in all known conformations of R.*Eco*RV[14]. In this 'open' free enzyme conformation R.*Pvu*II cannot contact DNA with both subunits as extensively as R.*Eco*RV (Fig. 4*c*). DNA can be contacted simultaneously by both monomers only in the region close to the local dyad (Fig. 4*c*) where the two hydrogen-bonded, local dyad-related loops are positioned; each loop connects helices αA and αB and comprises residues (Gln 33, Asp 34, Asn 35) which can form hydrogen bonds or salt bridges to DNA. Spatial alignment shows that Asp 34 in R.*Pvu*II is structurally equivalent to Lys 38 in R.*Eco*RV which is hydrogen bonded in the minor groove of the cognate DNA[14].

Structure–function relationships

The structure of the free enzyme form of R *Pvu*II allows several possible features of endonuclease structure–function relationships to be identified: a motif of secondary structural elements occurs in topologically equivalent positions in both known structures of endonucleases that cleave DNA leaving blunt ends (R.*Pvu*II and R.*Eco*RV). This structural homology is restricted to the DNA-binding subdomains, and no relation exists between dimerization subdomains. In the absence of sequence homology or extensive similarity in the DNA recognition sequences, the recurrence of structural features is probably related to basic requirements of common DNA binding and cleavage strategies. This hypothesis is supported by the fact that all DNA-binding regions of R.*Pvu*II are predictable on the basis of the topological equivalences to R.*Eco*RV discussed above, as ultimately confirmed by the recent structure determination of a R.*Pvu*II–DNA complex (X. Cheng, personal communication). Furthermore, enzymes that leave blunt ends are structurally nearly unrelated to R.*Eco*RI, an endonuclease that leaves 5' overhangs It is therefore possible that endonuclease structures form distinct classes, depending on the position of the scissile phosphates in DNA.

As already mentioned, the DNA-binding region close to the local dyad (residues Gln 33-Asn 35) is the only one in R.*Pvu*II where DNA can be contacted simultaneously by both monomers in the 'open' conformation of the free enzyme, if an analogous orientation between R.*Pvu*II and DNA is assumed as in the R.*Eco*RV/DNA complexes[14]. Early stages of DNA recognition by this 'open' conformation could thus involve the region close to the local dyad, and due to symmetry reasons, the central base pairs of cognate DNA. Because this DNA-binding region is positioned in the interface between subdomains A and B, it is possible that its interactions in the early steps of DNA recognition induce motions between the subdomains, which ultimately lead to a considerably more 'closed' conformation, capable of DNA binding. The tertiary structure changes of the protein may be thus associated with a stepwise recognition process of the DNA sequence, which starts at the central base pairs. The importance of subdomain flexibility for DNA binding has already been discussed in connection with R.*Eco*RV[14]. In the R.*Pvu*II sequence the potential for considerable flexibility between subdomains A and B is reflected in the clustering of three Gly residues between helices αA and αB and in the presence of a further Gly immediately after αB. The structural changes associated with the subdomain motions can be rather drastic: in the conformation of the R.*Pvu*II/DNA complex (X. Cheng, personal communication) residues His 85 of the two subunits share a proton, thus coming within hydrogen-bonding distance, while they are around 25 Å apart in the free enzyme. This indicates that the two DNA-

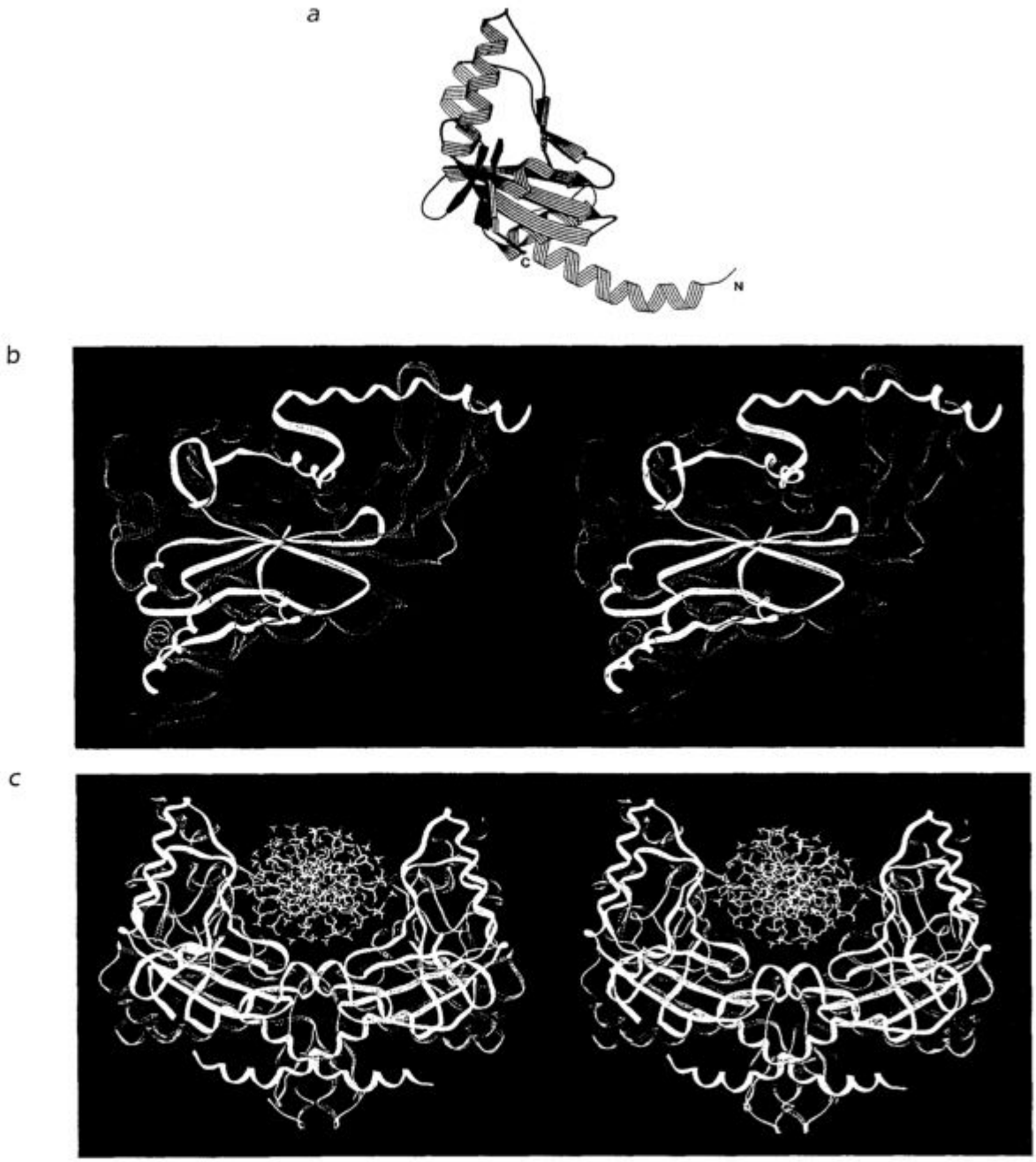


Fig. 4 Topological equivalences between R.*Pvu*II and R.*Eco*RV *a*, Schematic representation showing with light or dark shading the R.*Pvu*II segments that have topologically equivalent counterparts in R *Eco*RV Light shading indicates secondary structural elements that after spatial alignment of the two structures superimpose partially or completely with their equivalent counterparts in R.*Eco*RV with an r m.s.d. in the Cα positions better that 2 5 Å Dark shading indicates secondary structural elements that do not superimpose with their counterparts in R *Eco*RV *b*, Spatial alignment of R.*Pvu*II (yellow) with the free enzyme form of R.*Eco*RV (red); c, Spatial alignment of R.*Pvu*II dimer (yellow) with R.*Eco*RV (brown) complexed with noncognate DNA[14] Spatial alignments were performed using least-square techniques implemented in the program O[24] Best results (both for monomer and dimer alignments) are obtained when the transformations are performed between Cα atoms of αB, αD, αE, βc, βd, βe, βf, βg in R.*Pvu*II and their counterparts from αB, αD, αC, βe, βf, βg, βh, βj respectively in R.*Eco*RV

binding subdomains approach each other after the DNA is bound, whereby the cleft nearly closes with its width decreasing by more than 20 Å.

The structure of R.*Pvu*II suggests that extensive structural homologies may exist in endonucleases, despite the diversity of their sequences, and that these homologies may be related to functional properties. More structural studies including analyses of complexes with non-cognate DNA will be required in order to understand in detail the structure–function relationships of these important enzymes.

Methods

Purification and crystallization of R *Pvu*II have been reported earlier[20] The protein crystallizes in $P2_12_12$ with a=84 6, b=106 5 and c=47 0 (Å) with two monomers per asymmetric unit The crystal structure was determined by multiple isomorphous replacement–anomalous scattering (MIRAS) techniques using four derivatives (Table 1) X-ray intensities were collected using synchrotron radiation at the EMBL outstation at DESY (beamline X11, λ = 0 92 Å, MAR-Research imaging plate system), and on a Elliott GX21 rotating anode (CuK_α radiation, FAST area detector) Heavy atom sites were located in difference Patterson and difference Fourier maps and refined with the CCP4[21] program MLPHARE With the exception of one site of the Ag derivative (Table 1), all other heavy-atom sites are arranged in the dimer symmetrically relative to the local twofold axis The binding sites of the thiomersal[19] derivatives Hg and Hg1 (Table 1) are identical and formed in each monomer by Asn 35, Lys 38 and Ser 71 The heavy-atom sites of Pr and Yb (Table 1) are also nearly identical The lanthanides bind to the potential active-site residues Asp 58 and Glu 68, this could be of functional interest because it has been suggested[14] that in the R *Eco*RV structure the counterparts of these residues participate in the formation of the binding site of the essential Mg^{2+} ion All heavy-atom binding sites of the Ag derivative are formed by His residues

The position of the local dyad was determined from refined heavy-atom positions and confirmed results from self-rotation functions MIRAS phases calculated with MLPHARE after heavy atom refinement (figure of merit 0 59) were improved by solvent flattening[22] and molecular averaging Electron-density maps (Fig 1) were interpreted with the graphics programs FRODO[23] and O[24] The protein model was refined using XPLOR[25] initially by simulated annealing with the slow-cooling protocol[26] and at a later stage with conjugate gradient refinement (Table 1) Arg 54 and Glu 55 are disordered and were omitted from refinement Atomic coordinates have been deposited with the Brookhaven Data Bank

Note added in proof The recent structure determination of the *Bam*HI endonuclease[27] which reveals a considerable structural similarity to R *Eco*RI, supports our hypothesis that endonuclease structures can be classified on the basis of the position of the scissile phosphodiester bonds

Received 11 April; accepted 6 May 1994

Acknowledgements
We thank V Bouriotis and the Enzyme Technology Group of IMBB for technical support, and G Thireos, K Petratos and D Yphantis for reading the manuscript Prepublication access to the manuscript on the R *Pvu*II–DNA complex by X Cheng is acknowledged Data collection at the synchrotron radiation source was supported through a grant from the Large Installations Project Programme of the European Union

1. Wilson, G G & Murray, N E Restriction and modification systems *Rev Genet* **25**, 585–627 (1991)
2. Bennet, S P & Halford, S E Recognition of DNA by type II restriction enzymes *Curr Topics Cell Regul* **30**, 57–102 (1989)
3. Kruger, D H & Bickle, T A Bacteriophage survival multiple mechanisms for avoiding deoxyribonucleic acid restriction systems of their hosts *Microbiol Rev* **47**, 345–360 (1983)
4. Heitman, J How the *Eco*RI endonuclease recognizes and cleaves DNA *BioEssays* **14**, 445–454 (1992)
5 Roberts, R J & Macelis, D Restriction enzymes and their isoschizomers *Nucl Acids Res* **20**, 2167–2180 (1992)
6. Athanasiadis, A , Gregoriu, M , Thanos, D , Kokkinidis M , & Papamatheakis, J Complete nucleotide sequence of the *Pvu*II restriction enzyme gene from *Proteus vulgaris* *Nucl Acids Res* **18**, 6434 (1990)
7 Tao, T & Blumenthal, R M Sequence and characterization of the pvuIIR, the pvuII endonuclease gene, and of pvuIIC, its regulatory gene *J Bact* **174**, 3395–3398 (1992)
8. Gingeras, T R , Greenough, L , Schildkraut, I & Roberts R J Two new restriction endonucleases from *Proteus vulgaris* *Nucl Acids Res* **9**, 4525–4536 (1981)
9 Jack, W E *et al* Overexpression, purification and crystallization of *Bam*HI endonuclease *Nucl Acids Res* **19**, 1825–1829 (1991)
10. Wilson, G G Type II restriction-modification systems *Trends Genet* **4**, 314–318 (1988)
11. Kim, Y , Grable, J C , Love, R , Greene, P J & Rosenberg, J M Refinement of the *Eco*RI endonuclease crystal structure a revised protein chain tracing *Science* **249**, 1307–1309 (1990)
12. Rosenberg, J M Structure and function of restriction endonucleases *Curr Opin struct Biol* **1**, 104–113 (1991)
13. Winkler, F K Structure and function of restriction endonucleases *Curr Opin struct Biol* **2**, 93–99 (1992)
14. Winkler, F K *et al* The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments *EMBO J* **12**, 1781–1795 (1993)
15. Anderson, J E Restriction endonucleases and modification methylases *Curr Opin struct Biol* **3**, 24–30 (1993)
16. Priestle, J RIBBON: a stereo cartoon drawing program for proteins *J appl Crystallogr* **21**, 572–576 (1988)
17. Sänger, W in *Principles of Nucleic Acid Structure*, 386–389 (Springer Verlag, New York, 1988)
18. Fasman, G in *Prediction of Protein Structure and the Principles of Protein Conformation*, 193–316 (Plenum Press, New York, 1989)
19. Blundell, T L & Johnson, L N in *Protein Crystallography*, 205–219 (Academic Press, London, 1976)
20. Athanasiadis, A & Kokkinidis, M Purification, crystallization and preliminary X-ray diffraction studies of the *Pvu*II endonuclease *J molec Biol* **222**, 451–453 (1991)
21. CCP4, The SERC (UK) Collaborative Computing Project No 4, a Suite of Programs for Protein Crystallography (1979)
22. Wang, B C Resolution of phase ambiguity in macromolecular crystallography *Meth Enzymol* **115**, 90–112 (1985)
23. Jones, T A Interactive computer graphics FRODO *Meth Enzymol* **115**, 157–171 (1985)
24. Jones, T A , Zou, J Y , Cowan, S W & Kjeldgaard, M Improved methods for building protein models in electron density maps and the location of errors in these models *Acta Crystallogr* **A47**, 110–119 (1991)
25. Brünger, A T , Kuriyan, J & Karplus, M Crystallographic *R* factor refinement by molecular dynamics *Science* **235**, 458–460 (1987)
26. Brünger, A T , Krukowski, A & Erickson, J W Slow-cooling protocols for crystallographic refinement by simulated annealing *Acta Crystallogr* **A46**, 585–593 (1990)
27. Newman, M , Strzelecka, T , Dorner, L F , Schildkraut, I & Aggarwal, A K Structure of restriction endonuclease *Bam* HI and its relationship to *Eco* RI *Nature* **368**, 660–664 (1994)

Προσάρτημα ΙΙ

Καταχώριση της δομής της *Pvu*ΙΙ ενδονουκλεάσης στην Brookhaven βάση δεδομένων.

Προσάρτημα ΙΙΙ

Προγράμματα για τον προσδιορισμό της αλληλουχίας αναγνώρισης της "star" ειδικότητας της *Pvu*II ενδονουκλεάσης.

Το πρόγραμμα FindSequu δημιουργεί δεδομένης μίας αλληλουχίας ένα αρχείο που περιέχει όλα τα διαφορετικά εξανουκλεοτίδια της.

```
        PROGRAM FINDSEQUU

        CHARACTER*6 HEXA(5100),CUR
        INTEGER NOH,NOB,HEXAN(5100),I,POS
        CHARACTER*20 FILENAME
        CHARACTER*50 LINE1
        CHARACTER*50000 SEQU

        WRITE(6,*)'INPUT FILE NAME'
        READ(5,'(A)')FILENAME

        WRITE(6,*)'NUMBER OF BASES:'
        READ(5,*)NOB

        OPEN(1,FILE=FILENAME,STATUS='OLD')
        NOH=0
        POS=1
5       READ(1,10,END=20)LINE1
10      FORMAT(A)
        SEQU(POS:POS+50)=LINE1
        POS=POS+50
        GOTO 5
20      CLOSE(1)

        OPEN(1,FILE='LHEX.DAT',STATUS='NEW')
        POS=1
25      CUR=SEQU(POS:POS+5)
        DO I=1,NOH
        IF (HEXA(I).EQ.CUR) THEN
        HEXAN(I)=HEXAN(I)+1
        GOTO 30
        END IF
        END DO
        NOH=NOH+1
        HEXA(NOH)=CUR
        HEXAN(NOH)=1
30      POS=POS+1
        IF ((POS+6).EQ.NOB)GOTO 35
        GOTO 25

35      DO I=1,NOH
        WRITE(1,'(A)')HEXA(I)
        END DO

        END
```

Το πρόγραμμα Digest, με δεδομένα ένα αρχείο με αλληλουχίες εξανουκλεοτιδίων και μία DNA αλληλουχία, δημιουργεί τα πρότυπα πέψης για κάθε εξανουκλεοτίδιο, ελέγχει την συμφωνία τους με περιορισμούς του πειραματικού προτύπου και δημιουργεί ένα αρχείο με τα επιλεγμένα εξανουκλεοτίδια.

```
      PROGRAM DIGEST

      CHARACTER*6 CUR,PVUII,SEQU1,A1,B1,OCUR
      INTEGER NOH,NOB,  I,POS(3),POS1,FRAG(150),NFRAG,C1,C2,C3,K
      CHARACTER*20 FILENAME,OUTFILE,ENZFILE
      INTEGER C4,C5,C6,C7,C8,C9,CO,MIN
      CHARACTER*50 LINE1
      CHARACTER*50000 SEQU
      CHARACTER CIR
      REAL i1


      A1='ACGT'
      B1='TGCA'
      WRITE(6,*)'INPUT FILE NAME'
      READ(5,'(A)')FILENAME

      WRITE(6,*)'ENZYME FILE:'
      READ(5,'(A)')ENZFILE

      WRITE(6,*)'RESULT FILE:'
      READ(5,'(A)')OUTFILE

      WRITE(6,*)'IS IT A CIRCULAR MOLECULE ?'
      READ(5,'(A)')CIR

      PVUII='CAGCTG'

      OPEN(1,FILE=FILENAME,  STATUS='OLD',  CARRIAGECONTROL='LIST')
      POS1=1
5     READ(1,10,END=20)LINE1
10    FORMAT(A)
      SEQU(POS1:POS1+50)=LINE1
      POS1=POS1+50
      GOTO 5
20    CLOSE(1)

      CO=0
      OPEN(1,FILE=ENZFILE,STATUS='OLD')
      OPEN(2,FILE=OUTFILE,STATUS='NEW')
      I1=0
30    POS1=1
      I1=I1+1
      READ(1,'(A)',END=25),CUR
      SEQU1=CUR
      DO K=1,6
            OCUR(7-K:7-K)=B1(INDEX(A1,CUR(K:K)):INDEX(A1,CUR(K:K)))
      END DO
      NFRAG=1
35    POS(1)=INDEX(SEQU(POS1:),CUR)
      POS(2)=INDEX(SEQU(POS1:),PVUII)
      POS(3)=INDEX(SEQU(POS1:),OCUR)
      IF ((POS(1).EQ.0).AND.(POS(2).EQ.0).AND.(POS(3).EQ.0))GOTO 50
          IF (POS(1).LT.POS(2).AND.POS(1).NE.0)THEN
                          MIN=1
          ELSE
```

```
            IF(POS(2).EQ.0)THEN
                    MIN=1
            ELSE
                    MIN=2
            END IF
      END IF
      IF ((POS(MIN).GE.POS(3).OR.POS(MIN).EQ.0).AND.POS(3).NE.0)THEN
                     MIN=3
      END IF

      FRAG(NFRAG)=POS(MIN)
      NFRAG=NFRAG+1
      POS1=POS1+POS(MIN)
      GOTO 35
50    FRAG(NFRAG)=INDEX(SEQU(POS1:),' ')
      IF(CIR .EQ. 'Y' .OR. CIR .EQ.'y')THEN
           FRAG(1)=FRAG(1)+FRAG(NFRAG)
           NFRAG=NFRAG-1
      END IF
      C1=0
      DO I=1,NFRAG
            IF(FRAG(I).GT.6000.and.frag(i).lt.10000 )THEN
                    C1=C1+1
            END IF
            IF(FRAG(I).GT.1700 .AND. FRAG(I).LT.2300)THEN
                    C2=C2+1
            END IF

      END DO

      IF (C1.EQ.2.and.c2.ge.4)THEN
            CO=CO+1
            WRITE(2,'(A)')SEQU1
c           WRITE(6,*)CO
      END IF
      C1=0
      C2=0
      C3=0
      C4=0
      C5=0
      C6=0
      C7=0
      C8=8
      C9=0
      if(i1/100.eq.int(i1/100))then
      WRITE(6,*)I1,co
      end if
      GOTO 30


25    CLOSE(2)
      CLOSE(1)

      END
```

Απόδοση στην Ελληνική γλώσσα μή δόκιμων όρων της διεθνούς ορολογίας.

Area detector	Ανιχνευτής περιοχής
Blunt ends	Τυφλά άκρα
Conjugate gradient	Συνδιασμένη διαβάθμιση
Figure of merit	Δείκτης αξιοπιστίας
Hanging drop method	Μέθοδος κρεμαστής σταγόνας.
Image plate	Δίσκος ειδώλου.
Macro-seeding	Μάκρο-σπορά
Molecular Averaging	Μοριακή Ισοστάθμιση ή εξισορρόπιση
Multiple Isomorphous Replacement	Πολλαπλή Ισόμορφη Αντικατάσταση
Native protein	Φυσική πρωτεΐνη
Seeding	Σπορά
Self Rotation Function	Συνάρτηση αυτοπεριστροφής
Simulated Annealing	Προσομοιωμένη επαναδιάταξη
Slow-Cooling	Βραδεία Ψύξη
Solvent flattening	Εξομάλυνση πυκνότητας διαλύτη
Space group	Χωροομάδα
Overdigest	Υπερπέψη
Induced fit	Επαγώμενη προσαρμογή
Profile fitting	Ποσαρμογή με βάση την κατατομή.
Images	Είδωλα
Radiation damage	Φθορά ακτινοβολίας
Occupancy	Πληρότητα
Maximum Likelihood phase	Μέγιστης Πιθανότητας φάση
2-Fold Symmetry	Συμμετρία τάξεως-2
Loop	Βρόγχος
Lack of closure	Σφάλμα ανυσματικού αθροίσματος
Phasing Power	Ισχύς προσδιορισμού φάσεων
Backbone	Ραχοκοκκαλιά

ΣΥΝΤΜΗΣΕΙΣ ΚΑΙ ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

EMBL	Europian Molecular Biology Laboratory
DESY	Deutches Electronen SYnchrotron
LIP	Large Installation Project
EMBO	Europian Molecular Biology Organization
RMs	Restriction Modification systsem
ORF	Open Reading Frame
R.PvuII	Restriction enzyme PvuII
R.EcoRV	Restriction enzyme EcoRV
R.BamHI	Restriction enzyme BamHI
R.EcoRI	Restriction enzyme EcoRI
FOM	Figure Of Merit
SAM	S-Adenosyl-Methionine
HTH	Helix-Turn-Helix
HLH	Helix-Loop-Helix
MAD	Multiwavelength Anomalous Diffraction
CCM	Common Core Motif
DMSO	DiMethylSulfOxide
PC	PhosphoCellulose
GF	Gel Filtration
PEG	PolyEthylGlucose
AS	Ammonium Sulfate
R_{iso}	Risomorphous
K_{emp}	Kempirical
FFT	Fast Fourier Transform
PMSF	Phenyl Methyl Sylfonyl Fluoride
DTT	Dithiothreitol
E.coli	*Escherichia coli*
bp	base pairs
kb	kilo base pairs
PDB	Protein Data Bank

ΒΙΒΛΙΟΓΡΑΦΙΑ

Anderson J.E. (1993). **Restriction endonucleases and modification methylases.** *Current Opinion in Structural Biology.* **3,** 24-30.

Athanasiadis A. & Kokkinidis M. (1991). **Purification, crystallization and preliminary X-ray diffraction studies of the *Pvu*II endonuclease.** *J.Mol.Biol.* **222**, 451-453.

Athanasiadis A., Gregoriu M., Thanos D., Kokkinidis M. and Papamatheakis J. (1990). **Complete nucleotide sequence of the *Pvu*II restriction enzyme gene from *Proteus vulgaris.*** *Nucl. Acids Res.* **18**, 6434.

Athanasiadis A., Vlassi M., Kotsifaki D., Tucker P.A., Wilson K.S. and Kokkinidis M. (1994). **Crystal structure of *Pvu*II endonuclease reveals extensive structural homologies to *Eco*RV.** *Structural Biology* **1**, 469-475.

Ausubel F.M, Brent R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A. and Struhl K. (1987). **Current Protocols in Molecular Biology.** *Greene Publishing Associates and Wiley-Interscience.* ISBN 0-471-50338-X

Baker E.N. & Hubbard R.E.(1986). **Hydrogen Bonding in Globular Proteins.** *Prog. Biophys.Molec. Biol.* **44,** 97-179.

Balendiran K., Bonventre J., Knott R., Jack W., Benner J., Schildkraut I. and Anderson J.E. (1994). **Expression, purification and crystallization of restriction endonuclease *Pvu*II with DNA containing its recognition site.** *Proteins* **19**, 77-79

Barlett J.A., Gaillard R.K. Jr. and Joklik W.K. (1986). **Sequencing of supercoiled plasmid DNA.** *Biotechnicks* 4: 208-210.

Barlow D.J. & Thornton J.M. (1988). **Helix geometry in proteins.** *J. Mol. Biol.* **201,** 601-619.

Bennet S.P. & Halford S.E. (1989). **Recognition of DNA by type II restriction enzymes.** *Current Topics in Cellular Regulations* **30**, 57-79.

Biomol System Groningen (1991).

Birnboim H.C. (1983). **A rapid alkaline extraction method for the isolation of plasmid DNA.** *Methods in Ezymol.* **100**, 243-255.

Blow D.M. & Matthews B.W. (1973). **Parameter refinement in the multiple isomorphous-replacement method.** *Acta Cryst.* **A29**, 56-62.

Blow D.M. (1958). **The structure of hemoglobin.** *Proc. Roy. Sos.* **A247**, 302-336.

Blow D.M. (1985). **Introduction to rotation and translation functions.** *Proceedings of the Daresbury Study Weekend 15-16 February 1985*, 2-7.

Blumenthal R.M., Gregory S.A. and Cooperider J.S. (1985). **Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*.** *J. Bacteriol.* **164**, 501-509.

Blundell T.L. & Johnson L.N. (1976). **Protein crystallography.** Academic Press **ISBN: 0 12 108350-0.**

Bougueleret L., Schwarzstein M., Tsugita A. and Zabeau M. (1984). **Characterization of the genes for the *Eco*RV restriction and modification system of Escherichia coli.** *Nucleic Acids Research* **12**, 3659-3676.

Bradford M. (1976). **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of dye binding.** *Anal. Biochem.* **72**, 248-254.

Bricogne G. (1976). **Methods and Programs for Direct-Space Exploitation of Geometric Redundancies.** *Acta Crystallogr.* **A32**, 832.

Bricogne G. **Joy of Skewing**. Program write-up IBM version in CCP4 documentation.

Brooks J.E. (1987). **Properties and uses of restriction endonucleases.** *Methods Enzymol.* **152,** 113-129.

Brunger A.T. (1988). **Crystallographic Refinement by Simulated Annealing. Application to a 2.8 Å Resolution Structure of Aspartate Aminotransferase.** *J. Mol. Biol.* **203**, 803-816.

Brunger A.T. (1990). **X-PLOR Manual, v2.1.** (Yale University, New Haven, U.S.A.)

Brunger A.T., Krukowski A. and Erickson J. (1990). **Slow-cooling protocols for crystallographic refinement by simulated annealing.** *Acta crystallogr.* **A46**, 585-593.

Brunger A.T., Kuriyan J. and Karplus M. (1987). **Crystallographic R Factor Refinement by Molecular Dynamics.** *Science* **235,** 458-460.

Chandrasegaran S. & Smith H.O (1988). In Sarma M.H. and Sarma R.H. (eds). **Structure and Expression.** Adenine Press, Guilderland, N.Y. Vol **1**, 149-156.

Chen D., Liu Q., Chen X., Zhao X. and Chen Y. (1991). **The inhibition of restriction endonuclease PvuII cleavage activity by methylation outside its recognition sequence.** *Nucl. Acids Res.* **19**, 5703-5705.

Cheng X., Balendiran K., Schildkraut I., and Anderson J.E. (1994). **Structure of *Pvu*II Endonuclease with cognate DNA.** *EMBO Journal* **13**,3927-3935.

Cheng X., Kumar S., Posfai J., Pflugrath J.W. and Roberts R.J. (1993). **Crystal Structure of the *Hha*I DNA Methyltransferase complexed with S-adenosyl-L-methionine.** *Cell* **74**, 299-307.

Cull M. & McHenry C.S. (1990). **Preparation of extracts from Prokaryotes.** *Methods in enzymology,* **182**, 147-153.

Devereux J., Haeberli P. and Smithies O. (1984). **A comprehensive set of sequence analysis programs for the VAX.** *Nucleic Acids Res.* **12,** 9509-9524.

Dickerson E.R., Weinzierl J.E. and Palmer A.R. (1968). **A Least-Squares Refinement Method for Isomorphous Replacement.** *Acta Cryst.* **B24,** 997-1003.

Dodson E. & Vijayan M. (1971). **The Determination and Refinement of Heavy-Atom Parameters in Protein Heavy-Atom Derivatives. Some Model Calculations Using Acentric Reflections.** *Acta Cryst.* **B27**, 2402-2411.

Eliopoulos E.E., Geddes A.J., Brett M., Pappin D.J.C. and Findlay (1982). *J.B.C. Int. J. Biol. Macromol.* **4**, 263.

Ellenberger T. (1994) **Getting a grip on DNA recognition. Structures of the basic region leucine Zipper and the basic region helix-loop-helix DNA binding domains.** *Current Opinion in Structural Biology* **4**, 12-21.

Zabeau M. & Roberts R.J. (1979). **The role of restriction endonucleases in molecular genetics.** Academic Press Inc. ISBN 0-12-6844403-8.

Fox G.C. & Holmes K.C. (1966). **An Alternative Method of Solving the Layer Scaling Equations of Hamilton, Rollet and Sparks.** *Acta Cryst.* **20,** 886-891.

Frederick C.A.J., Melia M., Samudzi, Jen-Jacobson C. L., Wang B., Greene P., Boyer H.W. and Rosenberg J.M. (1984). **Kinked DNA in crystalline complex with *Eco*RI endonuclease.** *Nature* **309**,327-331.

Freemont S.P., Lane N.A. and Sanderson R.M. (1991). **Structural aspects of protein-DNA recognition.** *Biochem. J.* **278**, 1-23.

French S. & Wilson K. (1978). **On the treatment of Negative Intensity Observations.** *Acta Cryst.* **A34**, 517-525.

Gilliland G.L. & Davies D.R. (1990). **Protein Crystallization : The Growth of Large Scale Single Crystals.** *Methods in enzymology* **104**, 370-381.

Gingeras T.R., Greenough L., Schildkraut L. and Roberts R.J. (1981). **Two new restriction endonucleases from *Proteus vulgaris.*** *Nucl. Acids Res.* **9**, 4525-4536.

Gingeras T.R., Milazzo J.P. and Roberts R.J. (1978). **A computer assisted method for the determination of restriction enzyme recognition sites.** *Nucl. Acids Res.* **5,** 4105-4127.

Grasby A.J., & Connolly A.B. (1992). **Stereochemical Outcome of the Hydrolysis Reaction Catalyzed by the *Eco*RV Restriction Endonuclease.** *Biochemistry* **31**, 7855-7861.

Greer J. (1985). **Computer Skeletonization and Automatic Electron Density Map Analysis.** *Methods in enzymol.* **115,** 206-224.

Halazonatis D.T. & Kandil N.A. (1992). **Predicted Structural Similarities of the DNA Binding Domains of c-Myc and Endonuclease *Eco*RI.** *Science* **255**, 464-466.

Harker D. (1936). **The application of the three dimensional Patterson method and the crystal structures of prousite, Ag_3AsS_3, and pyragyrite Ag_3SbS_3.** *J. Chem. Phys.* **4**, 381-390.

Harrison S.C. & Aggarwal A.K. (1990). **DNA Recognition by Proteins with the Helix-Turn-Helix Motif.** *Annu. Rev. Biochem.* **59**, 933-969.

Heitman J. & Model P. (1990). **Substrate Recognition by the *Eco*RI Endonuclease.** *Proteins* **7,**185-197.

Hermans J. (1985) **Rationalization of Molecular Models.** *Methods in enzymol.* **115,** 171-189.

Heukeshoven J. & Dernick R. (1985). **Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining.** *Electrophoresis* **6**, 103-112.

Holmes D.S. & Quingley M. (1981). **A rapid boiling method for the preparation of bacterial plasmids.** *Anal. Biochem.* **119**, 93-95.

International Tables for X-ray Crystallography, vol. **A** (1989). Kluwer Academic Publishers.

Janin J. & Chothia C. (1985). **Domains in Proteins: Definition, Location and Structural Principles.** *Methods in enzymol.* **115,** 420-430.

Jeltsch A., Alves J., Maass G. and Pingoud A. (1992). **On the Catalytic Mechanism of the EcoRI and EcoRV.** *Febs lett.* **304,** 4-8.

Jones T.A. & Kjeldgaard M. (1992). **O - the Manual v.5.8.1.**

Jones T.A. & Thirup S. (1986). **Using Known Substructures in Protein Model Building and Crystallography.** *EMBO J.*, **5**, 819-822.

Jones T.A. (1988). **in FRODO: A graphics fitting program for macromolecule in computational crystallography.** ed. Sayre, D. 303-317 (Clarendon press, Oxford)

Jones T.A., Zou J.-Y. and Cowan S.W. (1991). **Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models.** *Acta Cryst.* **A47**, 110-119.

Jones Y. & Stuart D. (1991). **Locating heavy atom sites by automatic Patterson search - GROPAT.** *Isomorphous Replacement and Anomalous Scattering. Proceedings ot the CCP4 Study Weekend 25-26 January 1991*, 39-47.

Kabsch W. & Sander C. (1983). **Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features.** *Biopolymers* **22**, 2577-2637.

Kabsch W. (1988). **Automatic Indexing of Rotation Diffraction Patterns.** *J.Appl.Cryst.* **21**, 67-71.

Kalesnikov V.A., Zinoviev V.V., Yashina L.N., Karginov V.A., Baclacov M.M. and Malyguin E.G. (1981). **Relaxed Specificity of Endonuclease *Bam*HI as Determined by Identification of Recognition Sites in SV40 and pBR322 DNAs.** *FEBS Letters* **132**, 101-104.

Kim S-H. (1992). **β-Ribbon: A New DNA Recognition Motif.** Science 255, 1217-1218.

Kim Y.C., Grable J.C., Love R., Greene P.J. and Rosenberg J.M. (1990). **Refinement of the *Eco*RI endonuclease crystal structure: A revised protein chain tracing.** *Science* **249**, 1307-1309.

Klimasauskas S., Nelson L.J. and Roberts J.R. (1991). **The sequence specificity domain of cytosine-C5 methylases.** *Nucleic Acids Research* **19**,6183-6190.

Klimasauskas S., Timinskas A., Menkevicious S., Butkiene D., Butkus V. and Janulaitis A. (1989). *Nucl. Acids. Res.* **17**, 9823-9832.

Kruger D.H. and Bickle T.A. (1983). **Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts.** *Microbiol. Rev.* **47**, 345-360.

Kumar S., Cheng X., Pflugrath W.J., and Roberts J.R. (1992). **Purification, Crystallization, and Preliminary X-ray Diffraction Analysis of an M.HhaI-AdoMet Complex.** *Biochemistry* **31**, 8648-8653.

Lattman E. (1985). **Use of the Rotation and Translation Functions.** *Methods in enzymol.* **115,** 55-77.

Leslie A.G.W. (1991). **Heavy Atom Derivative Screening.** Isomorphous Replacement and Anomalous Scattering. *Proceedings ot the CCP4 Study Weekend 25-26 January 1991*, 9-22.

Leslie A.G.W., Brick P. and Wonacott A.J. (1986). *CCP4 Newsl.* **18,** 33-39.

Luzzati R. (1952). **Traitement Statistique des Erreurs dans la determination des Structures Cristallines.** *Acta Cryst.* **5**, 802-810.

MacArthur M.W. & Thornton J.M. (1991). **Influence of proline residues on protein conformation.** *J. Mol. Biol.* **218**, 397-412.

Mathews B.W. (1966). **The Determination of the Position of Anomalously Scattering Heavy Atom Groups in Protein Crystals.** *Acta Cryst.* **20**, 230.

Mathews B.W. (1968). **Solvent content of protein crystals.** *J. Mol. Biol.* **33**, 491-497.

Matthews B.W. & Rossman M.G. (1985). **Comparison of Protein Structures.** *Methods in enzymol.* **115,** 397-420.

McClarin J.A., Frederick C.A., Wang B., Greene P., Boyer W. and Rosenberg J.M. (1986). **Structure of the DNA-*Eco*RI Endonuclease Recognition Complex at 3 A Resolution.** *Science*, **234**,1526-1541.

McDonell M.W., Simon M.N. and Studier F.W. (1977). **Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels.** *J. Mol. Biol.* **110,** 119-146.

McPherson A. (1990). **Current approaches to macromolecular crystallization,** *Eur. J. Biochemistry* **189,**1-23.

McRee E.D. (1993). **Practical Protein Crystallography.** Academic Press, Inc. **ISBN 0-12-486050-8**

Moldrich P. & Roberts R.J. (1982). **Type II restriction and modification enzymes** p.109-154 in S.M. Linn and R.J. Roberts, Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Nakashima H., Nishikawa K. and Ooi T. (1986). **The Folding Type of a Protein is Relevant to the Amino Acid Composition.** *J. Biochem.* **99**, 153-162.

Nasri M. & Thomas D. (1986). **Relaxation of Recognition Sequence of Specific endonuclease *Hind*III.** *Nucl. Acids Res.* **14**, 811-821.

Nasri M. & Thomas D. (1987). **Alteration of the specificity of *Pvu*II restriction endonuclease.** *Nucleic Acids Research* **15**,7677-7687.

Nathans D. & Smith H.O. (1975). **Restriction Endonucleases in the Analysis and Restructuring of DNA Molecules.** *Annu. Rev. Biochem.* **44**, 273-293.

Newman M., Strzelecka T., Dorner L.F., Schildkraut I. and Aggarwal A.K. (1994b). **Structure of restriction endonuclease BamHI and its relationship to *Eco*RI.** *Nature* **368**,660-664.

Newman M., Strzelecka T., Dorner LF., Schildkraut I., and Aggarwal AK. (1994a). **Structure of Restriction Endonuclease *Bam*HI phased at 1.95A resolution by MAD analysis.** *Structure* **2**, 5, 439-452.

Oakley B.R., Kirsch D.R. and Morris N.R. (1980). **A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels.** *Anal. Biochem.* **105**, 361-363.

Otwinowski Z. (1991). **A Maximum-Likelihood Theory of Heavy Atom Parameter Refinement in the Isomorphous Replacement Method.** *Proceedings ot the CCP4 Study Weekend 25-26 January 1991*, 80-85.

Otwinowski, Sbyszek. (1993). **Maximum likelihood refinement of heavy atom parameters.** In: (Wolf et al., 1991). DL/SCI/R32, ISSN 0144-5677.

Patterson A.L. (1934). **A Fourier series method for the determination of the components of interatomic distances in crystals.** *Phys. Rev.* **46**, 372-376.

Perutz M.F. (1956). **Isomorphous Replacement and Phase Determination in Non-centrosymmetric Space Groups.** *Acta Cryst.* **9,** 867-.

Polisky B., Greene P., Garfin D.E., Mc Carthy B.J., Goodman H.M. and Boyer H.W. (1975). **Specificity of Substrate Recognition by the EcoRI Restriction Endonuclease.** *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3310-3314.

Posfai S., Bhagwat A.J., Posfai G. and Roberts R.J. (1989). **Predictive Motifs Derived from Cytosine Methyltransferases.** *Nucl. Acids Res.* **17**,2471-2435.

Powell M.J.D. (1977). *Mathematical Programming* **12,** 241-254.

Priestle P.J. (1988). **RIBBON: a stereo cartoon drawing program for proteins.** *J. Appl. Cryst.* **21**, 572-576.

Radloff R., Bauer W. and Vinograd J. (1967). **A dye-buoyant-density method for the detection and isolation of closed circular DNA in HELA cells.** *Proc. Natl. Acad. Sci. USA* **57**, 1514-1521.

Rainer J. & Rainer R. (1989). **Folding Proteins.** (in Protein Structure a practical approach). IRL Press at Oxford University Press. ISBN 1 85221 137 7.

Ramachandran G.N. & Sasisekharan V. (1968). **Conformation of Polypeptides and Proteins.** *Adv. Protein Chem.* **23**, 283-437.

Read R.J. (1986). **Improved Fourier coefficients for map using phases from partial structures with errors.** *Acta crystallogr.* **A46**, 585-593.

Reich O.N. & Danzitz J.M. Jr. (1992). ***Eco*RI DNA Methyltransferase-DNA Interactions.** *Biochemistry* **31**,1937-1945.

Reich O.N., & Danzitz Jr,J.M. (1991). **Non-additivity of sequence-specific enzyme-DNA interactions in the *Eco*RI DNA methyltransferase.** *Nucleic Acids Research* **19**,6587-6594.

Richardson J.S. & Richardson C.D. (1985). **Interpretation of Electron Density Maps.** *Methods in enzymol.* **115,** 189-206.

Richardson S.J. (1985). **Describing Patterns of Protein Structure.** *Methods in enzymol.* **115,** 341-358.

Roberts R.J. & Halford S.E. (1993). **Type II restriction endonucleases. In Nucleases.** 2nd edn. Cold Spring Harbor, Laboratory Press, New York.

Roberts R.J. & Macelis D. (1992). **Restriction enzymes and their isoschizomers.** *Nucleic Acids Res.* **20,** 2167-2180.

Sambrook J., Fritch E.F. and Maniatis T. (1989). **Molecular cloning: a laboratory manual.** (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

Sanger F. & Coulson A.R. (1978). **The use of thin acrylamide gels for DNA sequencing.** *FEBS lett.* **87**, 107-110.

Sanger F., Nicklen S. and Coulson A.R. (1977). **DNA sequencing with chain terminating inhibitors.** *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Schmiedeskamp M. & Klevit R.E. (1994). **Zinc finger diversity.** *Current opinion in Structural Biology* **4**, 28-35.

Selent U., Ruter T., Kohler E., Michaela L., Thielking V., Alves J., Oelgeschlager T., Wolfes H., Peters F. and Pingoud A. (1992). **A Site-Directed Mutagenesis Study To Identify Amino Acid Residues Involved in the Catalytic**

Function of the Restriction Endonuclease *Eco*RV. *Biochemistry* **31**,4808-4815.

SERC Daresbury Laboratory (1979). **CCP4. A Suite of Programs for Protein Crystallography.** Daresbury Laboratory, Warrington, England.

Sharp P.A., Sugden B. and Sambrook J. (1973). **Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose.** *Biochemistry* **12**, 3055-3063.

Smith M.D., Longo M., Gerard G.F. and Chatterjee D.K. (1992). **Cloning and characterization of genes for the *Pvu*I restriction and modification system.** *Nucleic Acids Research* **20**,21,5743-5747.

Sparks R.A. (1985). **Least-Squares Refinement.** *Methods in enzymol.* **115,** 23-41

Spolar S.R. & Record Jr. M.T. (1994). **Coupling of Local Folding to Site-Specific Binding of Proteins to DNA.** *Science* **263**, 777-784.

Stephenson F.H., Ballard B.T., Boyer H.W., Rosenberg J.M. and Greene P.S. (1989) **Comparison of the nucleotide and amino acid sequences of the *Rsr*I and *Eco*RI restriction endonucleases.** *Gene,* **85**, 1-13.

Tao T. & Blumenthal R.M. (1992). **Sequence and characterization of *pvuIIR*, the *PvuII* endonuclease gene, and of *pvuIIC*, its regulatory gene.** *Journal of Bacteriology,* 174, 10, 3395-3398.

Tao T., Bourne J.C. and Blumenthal R.M. (1991). **A family of regulatory genes associated with type II restriction-modification systems.** *J. Bacteriol.* **173**, 1367-1375.

Tao T., Walter J., Brennan K.J., Cotterman M.M. and Blumenthal R.M. (1989). **Sequence, internal homology and high-level expression of the gene for a DNA-(cytosine N4)-methyltransferase, M.*Pvu*II.** *Nucleic Acid Research* 17,**11**,4161-4175.

Taylor D.J., Badcoe I.G., Clarke A.R. and Halford S.E. (1991). ***Eco*R V Restriction Endonuclease Binds All DNA Sequences with Equal Affinity.** *Biochemistry* 30, **36,** 8743-8752.

Ten Eyck L. (1973). **Crystallographic Fast Fourier Transforms.** *Acta Cryst.* **A29**, 183-191.

Ten Eyck L. (1976). **Efficient Structure-Factor Calculation for Large Molecules by the Fast Fourier Transform.** *Acta Cryst.* **A33**, 486-492.

Ten Eyck L. (1985). **Fast Fourier Transform Calculation of Electron Density Maps.** *Methods in enzymol.* **115,** 324-337.

Thielking V., Selent U., Kohler E., Landgraf A., Wolfes H., Alves J. and Pingoud A. (1992). **$Mg2^+$ Confers DNA Binding Specificity to the *Eco*R V Restriction Endonuclease.** *Biochemistry* **31,** 3727-3732.

Thielking V., Selent U., Kohler E., Wolfes H., Pieper U., Geiger R., Urbanke C., Winkler K.F., and Pingoud A. (1991). **Site-Directed Mutagenesis Studies with *Eco*RV Restriction Endonuclease To Identify Regions Involved in Recognition and Catalysis.** *Biochemistry* **30**, 6416-6422.

TRIPOS Assosiates Inc. (1992). **SYBYL Molecular Modeling Software V 6.0.** *Manual.*

Tronrud D.E., Ten Eyck L.F. and Matthews B.W. (1987). **An efficient general-purpose least-squares refinement program for macromolecular structures.** *Acta crystallogr* **A43**, 489-501.

Vermote L.M.C. & Halford E.S. (1992). ***Eco*RV Restriction Endonuclease : Communication between Catalytic Metal Ions and DNA Recognition.** *Biochemistry* **31**,6082-6089.

Vermote L.M.C., Vipond I.B. and Halford E.S. (1992). ***Eco*RV Restriction Endonuclease : Communication between DNA Recognition and Catalysis.** *Biochemistry* **31**, 6089-6097.

Wang B.C (1985a). **Resolution of Phase Ambiguity in Macromolecular Crystallography.** *Methods in Enzymology* **115,** 90-112.

Wang B.C. (1985b). **in Diffraction Methods for Biological Macromolecules** (Wyckoff, H., ed.) Academic Press, New York.

Watenpaugh D.K. (1985). **Overview of Phasing by Isomorphous Replacement.** *Methods in enzymol.* **115,** 3-15.

Wilson G.G (1991). **Organization of restriction-modification systems.** *Nucleic Acids Research* **19**, 10, 2539 - 2566.

Wilson G.G. & Murray N.E. (1991). **Restriction and Modification Systems.** *Annu. Rev. Genet.* **25,** 585-627.

Wilson, G.G. (1988). **Type II restriction-modification systems,** *Trends in Genetics* **4**, 314-318.

Winkler F.K. (1992). **Structure and function of restriction endonucleases.** *Current opinion in Structural Biology* **2**,93-99.

Winkler F.K., Banner D.W., Oefner C., Tsernoglou D., Brown R.S., Heathman S.P., Bryan, R.K., Martin P.D., Petratos K. and Wilson K. S. (1993). **The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments.** *EMBO Journal* **12**,1781-1795.