ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ
ΙΝΣΤΙΤΟΥΤΟ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΟΥ ΤΟΥ ΣΥΜΠΛΕΓΜΑΤΟΣ ΤΗΣ ΠΡΩΤΕΙΝΗΣ CF2 ΚΑΙ DNA ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΑΛΛΗΛΟΥΧΙΩΝ ΔΑΚΤΥΛΩΝ ΨΕΥΔΑΡΓΥΡΟΥ

Δ.Ν. ΠΑΡΑΣΚΕΥΗΣ

ΔΙΑΤΡΙΒΗ ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΕΙΔΙΚΕΥΣΗΣ

Φεβρουάριος 1994

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τους δρ. Μ. Κοκκινίδη και τον ερευνητή κ. Κ. Πετράτο για την μεγάλη βοήθεια και την αμέριστη συμπαράσταση τους κατά την διάρκεια αυτής της εργασίας.

Τον συνεργάτη Κ. Παλιακάση για την βοήθεια του στην ανάπτυξη των προγραμμάτων και τα εποικοδομητικά σχόλια κατά την συγγραφή. Τα μέλη της ομάδας κρυσταλλογραφίας πρωτεϊνών δρ. Μ. Βλάσση, Α. Αθανασιάδη Ι. Παπανικολάου, και Σ. Τριανταφυλλάκη για τις χρήσιμες παρατηρήσεις τους. Επίσης τους Μ. Πιτταροκοίλη, και τις Ν. Κοτσιφάκη, και Ε. Βακουφάρη που χωρίς την βοήθεια τους θα ήταν αδύνατον να πραγματοποιηθεί η παρούσα εργασία.

Θα ήθελα να ευχαριστήσω επίσης τους δρ. Μ. Τσαγρή, και δρ. Β. Μπουριώτη για την βοήθεια τους κατά την διάρκεια της συγγραφής αυτής της εργασίας.

Η εργασία αυτή χρηματοδοτήθηκε από προγράμματα του Ινστιτούτου Μοριακής Βιολογίας και Βιοτεχνολογίας.

ΠΕΡΙΕΧΟΜΕΝΑ

Ευχαριστίες 1
Περιεχόμενα 2
Περίληψη 4
Summary 6

ΚΕΦΑΛΑΙΟ Ι **ΕΙΣΑΓΩΓΗ**

Ι. Εισαγωγή
Γενικά 8
Ι.Ι Πρωτεΐνες που αλληλεπιδρούν με DNA 10
Έλικα-στροφή-έλικα (HTH) 10
Φερμουάρ λευκίνης (bzip) 11
β κορδέλα 11
Ι.ΙΙ Πρωτεΐνες που δεσμεύουν Zn^{+2} 12
Ι.ΙΙΙ Βιολογική λειτουργία της πρωτεΐνης CF2 18
Ι.ΙV Λόγοι πραγματοποίησης αυτής της εργασίας 19

ΚΕΦΑΛΑΙΟ ΙΙ **ΠΕΙΡΑΜΑΤΙΚΑ**

ΙΙ. Μοντέλο του CF2 με το DNA και στατιστική ανάλυση αλληλουχιών δακτύλων Zn
ΙΙ.Ι Γενικά 23
ΙΙ.ΙΙ ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ 23
ΙΙ.ΙΙ.1 Μοντέλο CF2 με το DNA 23
ΙΙ.ΙΙ.2 Συλλογή και στατιστική ανάλυση των αλληλουχιών δακτύλων ψευδαργύρου 26
ΙΙ.ΙΙ. 3 Μεθοδολογία 27

ΚΕΦΑΛΑΙΟ ΙΙΙ **ΑΠΟΤΕΛΕΣΜΑΤΑ**

ΙΙΙ. Αποτελέσματα
ΙΙΙ.Ι Μοντέλο του συμπλέγματος του CF2 με το DNA 31
ΙΙΙ.ΙΙ Στατιστική ανάλυση 38
ΙΙΙ.ΙΙ.1 Στατιστική των αμινοξέων που αλληλεπιδρούν

ειδικά με το DNA 39
III.II.2 Στατιστική για τις υπόλοιπες θέσεις της έλικας 43
III.II.3 Στατιστική για την υπόλοιπη αλληλουχία
των δακτύλων ψευδαργύρου 48
III.II.4 Στατιστική για τις υπόλοιπες οικογένειες
δακτύλων ψευδαργύρου 54

ΚΕΦΑΛΑΙΟ IV **ΣΥΜΠΕΡΑΣΜΑΤΑ**

VI. ΣΥΖΗΤΗΣΗ ΚΑΙ ΜΕΛΛΟΝΤΙΚΑ ΣΧΕΔΙΑ
VI.I Συζήτηση 56
VI.II Μελλοντικά σχέδια 57

ΠΑΡΑΡΤΗΜΑ

Πίνακας 2. 60
Πίνακας 7. 64
Πίνακας 9. 66
Πίνακας 10. 67
Πίνακας 11 68
Προγράμματα που αναπτύχθηκαν για αυτήν την
εργασία 85

ΠΕΡΙΛΗΨΗ

Οι δάκτυλοι ψευδαργύρου (zinc fingers) είναι μικρές περιοχές (domains) αποτελούμενες από περίπου 28 αμινοξέα, που αναγνωρίζουν επιλεκτικά συγκεκριμένες αλληλουχίες DNA. Η παρούσα εργασία αναφέρεται i) στην δημιουργιά μοντέλου του συμπλέγματος τριών δακτύλων ψευδαργύρου της πρωτεΐνης CF2 (chorion factor-2) από δροσόφιλα με το DNA και ii) στη στατιστική ανάλυση 577 τέτοιων αλληλουχιών από 106 διαφορετικές πρωτεΐνες δακτύλων ψευδαργύρου.

Το μοντέλο του συμπλόκου της πρωτεΐνης CF2 με το DNA στηρίχτηκε αποκλειστικά στις συντεταγμένες της λυμένης δομής του συμλόκου Zif268 με το DNA (Pavletich and Pabo, 1991). Η πρωτεΐνη Zif268 αποτελεί χαρακτηριστικό παράδειγμα των αλληλεπιδράσεων μεταξύ πρωτεϊνών δακτύλου ψευδαργύρου και του DNA.

Η πρωτεΐνη CF2 (chorion transcription factor) αποτελεί ρυθμιστικό παράγοντα γονιδίων στο χώριον (chorion) της δροσόφιλα (Drosophila melanogaster). Ο λόγος πραγματοποίησης του μοντέλου ήταν κυρίως για να ερευνηθεί αν τηρούνται οι ίδιοι κανόνες αλληλεπίδρασης με το DNA, όπως αυτοί περιγράφτηκαν σε γνωστές δομές πρωτεϊνών του είδους (Pavletich and Pabo, 1991). Η αλληλουχία που αναγνωρίζεται από την πρωτεΐνη CF2 περιέχει κυρίως ζευγάρια βάσεων A/T (5'-GATTATATA-3'), σε αντίθεση με αυτή που αναγνωρίζεται από την πρωτεΐνη Zif268 που περιέχει κυρίως ζεύγη G/C (5'-GCGTGGGCG-3') (Pavletich and Pabo, 1991).

Τα αποτελέσματα έδειξαν ότι και στο μοντέλο του CF2 τα γενικά χαρακτηριστικά της αλληλεπίδρασης διατηρούνται όπως και στο μοντέλο του Zif268 (Pavletich and Pabo, 1991). Επιπλέον όμως παρατηρήθηκαν και κάποια καινούργια στοιχεία, όπως η δυνατότητα αλληλεπίδρασης με βάσεις από το δευτερεύοντα κλώνο (secondary strand) του DNA, ή η δυνατότητα σχετικής τοποθέτησης των δακτύλων

ψευδαργύρου, προς την κύρια αύλακα ανάλογα με το μήκος των πλευρικών αλυσίδων στις κρίσιμες θέσεις.

Από την στατιστική ανάλυση των αλληλουχιών δακτύλων ψευδαργύρου προέκυψαν χρήσιμα συμπεράσματα για τον ρόλο που παίζει κάθε θέση στη διεργασία αναγνώρισης του DNA. Επίσης βρέθηκαν ποιά συγκεκριμένα αμινοξέα και ποιοί συνδυασμοί τους απαντούν συχνώτερα στις κρίσιμες (αυτές δηλαδή που αλληλεπιδρούν με τις βάσεις του DNA) ή μη θέσεις ενός δακτύλου ψευδαργύρου.

Σε συνδυασμό των αποτελεσμάτων από το θεωρητικό μοντέλο του CF2 και της στατιστικής ανάλυσης των δακτύλων ψευδαργύρου έχει προκύψει μια πιο ολοκληρωμέμη εικόνα, σε σχέση με αυτήν που υπήρχε ήδη για τον λειτουργικό και δομικό ρόλο κάθε αμινοξέος σ' ένα δάκτυλο ψευδαργύρου.

Summary

The zinc finger is a small independently folded DNA recognition motif found in many eykaryotic proteins. In order to analyze them a database of 577 zinc fingers, from 106 different proteins has been constructed and a program for statistical analysis of aligned sequences written. We also designed a three dimensional model for the CF2-DNA interaction based on the structure of the Zif268-DNA complex (Pavletich and Pabo, 1991). The previous one represents an excellent working model for the interactions between the zinc fingers and the DNA.

The chorion transcription factor CF2 was isolated as a potentional transcriptional regulator of a chorion gene of Drosophila melanogaster. It is a zinc finger protein which is present in the nuceli of follicle cells which produce the chorion. Localized CF2 is not detected in the early embryo but appears during later embryonic stages (Hsu et al., 1992). The model of the CF2-DNA complex was of general interest because it predicts the modularity of finger-DNA interactions other than the known hydrogen bonding has been reported at the Zif268-DNA structure. The binding site of the CF2 protein is A/T rich DNA (5'-GATTATATA-3') instead of the G/C rich DNA (5'-GCGTGGGCG-3') of the Zif268 protein.

The results showed apart of the well characterized pattern observed in the Zif268 structure, that additional specific contacts may be formed with bases of the secondary DNA strand.

From statstical analysis of the zinc finger sequences the particular role of each amino acid was noted and clusters of correlated residues at critical positions (DNA base contact positions) were observed. The results of the theoretical model of the complex CF2- DNA cmbined with statistical analysis will help to understand the mechanism of how zinc fingers, where the structure is unknown interact with the DNA.

ΚΕΦΑΛΑΙΟ Ι

ΕΙΣΑΓΩΓΗ

I. ΕΙΣΑΓΩΓΗ

Γενικά

Ένα από τα πιο ενδιαφέροντα προβλήματα στην βιολογία σήμερα, είναι το πώς ενεργοποιούνται τα γονίδια σε πολυκύτταρους οργανισμούς. Για να ενεργοποιηθεί ένα γονίδιο μια σειρά πρωτεϊνών, γνωστοί ως μεταγραφικοί παράγοντες (*transcription factors*) πρέπει να αλληλεπιδράσουν με ένα μέρος του γονιδίου, τον υποκινητή (*promoter*). Αυτός ο μηχανισμός αποτελεί ένα είδος διακόπτη και επιτρέπει σε μια γενετική πληροφορία να μεταγραφεί από DNA σε RNA. Το ερώτημα είναι, πώς είναι δυνατόν ένας μεταγραφικός παράγοντας να επιλέξει ως στόχο την συγκεκριμένη αλληλουχία του υποκινητή, ανάμεσα σε πολύ μεγάλο αριθμό βάσεων;

Πολλοί μεταγραφικοί παράγοντες περιέχουν μικρές περιοχές (*domains*) που ονομάζονται δάκτυλοι ψευδαργύρου (*zinc fingers*) και έχουν την ικανότητα να επιλέγουν με μεγάλη ακρίβεια συγκεκριμένες αλληλουχίες DNA. Πρωτεΐνες αυτού του είδους ανακαλύφθηκαν για πρώτη φορά το 1985 σε βάτραχο (*Freemont et al., 1991*), και από τότε έχουν αναγνωριστεί γύρω στις διακόσιες με παρόμοια λειτουργία.

Είναι πολύ σημαντικό να κατανοήσουμε σε μοριακό επίπεδο αυτόν τον μηχανισμό, που απαιτεί γνώση της τρισδιάστατης διαμόρφωσης των πρωτεϊνών και των συμπλόκων τους με το DNA. Τα τελευταία χρόνια πειραματικές μελέτες κρυσταλλογραφίας ακτίνων-Χ όσο και φάσματα πυρηνικού μαγνητικού συντονισμού N.M.R., σε μόρια πρωτεϊνών και σε σύμπλοκα τους με DNA έχουν βοηθήσει σημαντικά την έρευνα στο παραπάνω πεδίο.

Υπάρχουν αρκετές ομάδες πρωτεϊνών που κατατάσσονται βάσει της ομολογίας τους σε επίπεδο πρωτοταγούς δομής. Οι παραπάνω ομάδες φαίνονται στον Πίνακα 1 μαζί με παραδείγματα πρωτεϊνών που η δομή τους έχει προσδιοριστεί πειραματικά. Τα είδη που αναφέρονται παρουσιάζουν πολυμορφία σχετικά με τον τρόπο δράσης τους. Οι επαφές πραγματοποιούνται συνήθως από έλικες δια μέσω δεσμών υδρογόνου, και μη πολικών *van der Waals* αλληλεπιδράσεων. Στις περισσότερες περιπτώσεις τα μόρια τοποθετούνται στην κύρια αύλακα (*major groove*) του DNA. Επίσης οι αλληλεπιδράσεις με τα άτομα του σκελετού (*backbone*) του DNA, πιο συγκεκριμένα δεσμοί υδρογόνου με τα οξυγόνα του φωσφόρου, είναι εξίσου σημαντικές για τον σωστό προσανατολισμό της πρωτεΐνης ως προς τις βάσεις. Και αυτές λοιπόν αποτελούν αναμφισβήτητα, σημαντικό τμήμα της διεργασίας αναγνώρισης της αλληλουχίας στόχου.

ΠΙΝΑΚΑΣ 1 Πρωτεΐνες που αλληλεπιδρούν με DNA και έχουν λυθεί πειραματικά

Πρωτεΐνη	Πειραματικά δεδομένα	Αλλοστερικά φαινόμενα	Αναφορά
Ομάδα I (HTH)			
Ρυθμιστικές πρωτεΐνες ευκαρυωτικών οργ. Homedomains			
Antp	NMR	OXI	Qian & Otiing (1989)
engrailed	X-ray	OXI	Kissinger et al. (1990)
Ρυθμιστικές πρωτεΐνες προκαρυωτικών οργ.			
lac καταστολέας	NMR	λακτόζη	Kaptein et al. (1985)
λ καταστολέας	X-ray	OXI	Jordan & Pabo (1982)
434 καταστολέας	X-ray	OXI	Mondragon et al. (1989)
434 cro	X-ray	OXI	Mondragon et al. (1989)
λ cro	X-ray	OXI	Brennan et al (1990)
καταβολικός ενεργοποιητής (CAP)	X-ray	c-AMP	Schuultz et al (1991)
καταστολέας της trp	X-ray	L-Trp	Otwinowski et al (1988)
inversion stimulation factor (FIS)	X-ray	OXI	Kostrewa et al (1991)
Ομάδα II (Πρωτεΐνες που δεσμεύουν Zn)			
Τάξη 1 (Δάκτυλοι Zn)			
fin (δάκτυλος 31)	NMR	OXI	Lee et al (1989)
ADR1 (δάκτυλος ADR1b)	NMR	OXI	Klevit et al (1990)
Zif 268 (δάκτυλοι 1-3)	X-ray	OXI	Pavlevitc & Pabo (1991)
Τάξη 2 (LZnH)			
υποδοχέας γλυκοκορτκοειδούς	NMR; X-ray	γλυκοκορτικοειδές	Hard & Luisi (1990)
υποδοχέας οιστρογόνου	NMR	οιστρογόνο	Schwabe et al (1990)
Ομάδα III (Φερμουάρ λευκίνης)			
GCN4	NMR	OXI	Oas et al (1990)
Ομάδα IV (β κορδέλα)			
Τάξη 1			
Met J	X-ray	S-αδενυλ-μεθειονίνη	Philips et al (1989)
Arc	NMR	OXI	Berg et al (1990)
Τάξη 2			
HU	X-ray	OXI	White et al (1989)

Ι.Ι Πρωτεΐνες που αλληλεπιδρούν με DNA

Έχουμε κατατάξει τις παραπάνω πρωτεΐνες, όπως φαίνεται στον Πίνακα 1 σύμφωνα με δομικά χαρακτηριστικά και σε υποκατηγορίες ανάλογα με επιμέρους κριτήρια. Κατατάσουμε λοιπόν τις πρωτεΐνες σε τέσσερις μεγάλες ομάδες και αναφέρουμε παραδείγματα για αυτές που υπάρχουν δομικά δεδομένα.

Έλικα-στροφή-έλικα (HTH)

Από τις πρώτες κρυσταλλογραφικές εργασίες στον λ καταστολέα έγινε φανερό ότι οι συγκεκριμένες πρωτεΐνες αλληλεπιδρούν με το DNA μέσω δύο ελίκων. Αυτό το πρότυπο χαρακτηρίζεται σήμερα ως έλικα-στροφή-έλικα (*helix-turn-helix*) και περιλαμβάνει μεταξύ των ελίκων, που σχηματίζουν μεταξύ τους μια γωνία περίπου 90^{o}, μια β-στροφή. Μια από τις έλικες (Εικ.1) τοποθετείται στην κύρια αύλακα, και είναι κυρίως υπεύθυνη για την επιλεκτική αλληλεπίδραση (*sequence specific*) με το DNA. Το μοτίβο (HTH) έχει βρεθεί σε αρκετούς προκαρυωτικούς καταστολείς και ενεργοποιητές, όπως φαίνεται στον Πίνακα 1. Τελευταία έχουν βρεθεί παρόμοια πρότυπα και σε πρωτεΐνες *homeodomain*, για παράδειγμα στην δροσόφιλα έχει βρεθεί η πρωτεΐνη *antennapedia homeodomain*. Σε μερικές πρωτεΐνες αυτής της ομάδας (δες Πίνακα 1) παρατηρούνται αλλοστερικά φαινόμενα όπως για την πρωτεΐνη *CAP* (*Schultz et al.*, 1990) που απαιτείται δηλαδή και η παρουσία ενός επιπλέον μορίου (*c-AMP*) για να λειτουργήσει η συγκεκριμένη πρωτεΐνη.

Φερμουάρ λευκίνης (β Zip)

Στις πρωτεΐνες αυτές το αμινοτελικό τους άκρο αποτελεί την βασική περιοχή της πρωτεΐνης (*basic region*) που πιστεύεται ότι είναι κυρίως υπεύθυνη για τις αλληλεπιδράσεις με το DNA. Η συγκεκριμένη περιοχή είναι θετικά φορτισμένη και αποτελείται κυρίως από α-έλικα που μαζί με την δεύτερη υπομονάδα σχηματίζουν το δομικό πρότυπο της υπερέλικας (*coiled coil*) (*Oas et al.*, 1990). Παρόμοια με άλλα τέτοια πρότυπα και το συγκεκριμένο παρουσιάζει μια περιοδικότητα στην κατανομή ανά επτά αμινοξέα. Εμφανίζονται δηλαδή στην πρώτη και στην τέταρτη θέση της επτάδας, που βρίσκονται στην ίδια πλευρά λόγω της α-έλικας, μη πολικές (*non polar*) πλευρικές αλυσίδες. Για παράδειγμα στην πρωτεΐνη του μύκητα *GCN4* η τέταρτη θέση καταλάμβανεται σε ποσοστό 80% από λευκίνη (*Landschulz*, 1988). Το λειτουργικό μέρος του μορίου περιλαμβάνει περίπου τριάντα, ως επί το πλείστον βασικά αμινοξέα που διπλώνουν κυρίως σαν α-έλικα.

β κορδέλα (β ribbon)

Υπάρχουν δύο τάξεις πρωτεϊνών σε προκαρυωτικούς οργανισμούς που αλληλεπιδρούν με το DNA δια μέσω της διαμόρφωσης της β κορδέλας (Εικ. 2). Η πρώτη περιλαμβάνει τον καταστολέα *Met J* από *Escherichia coli*, και επίσης τους καταστολείς *arc* και *mnt* από *Salmonella phage P22*. Η δεύτερη τάξη περιλαμβάνει την πρωτεΐνη *HU* που σχηματίζει νουκλεοπρωτεϊνικό σύμπλεγμα (*nucleoprotein structures*) σε πολλούς προκαρυωτικούς οργανισμούς. Η δομή του *Met J* έχει προσδιοριστεί κρυσταλλογραφικά (*Phillips*, 1991 - *Raffery et al.*, 1989), και η δομή του *Arc* με NMR (*Berg et al.*, 1990). Αυτές οι πρωτεΐνες σχηματίζουν διμερή με ένα πυρήνα από τέσσερις α-έλικες, δύο από κάθε υπομονάδα. Επίσης σχηματίζεται ένα πρότυπο β κορδέλας στο αμινοτελικό άκρο, εξέχοντας από τον πυρήνα. (Εικ. 8) Στο σύμπλοκο του *Met J* με το DNA φαίνεται ότι αυτή η β κορδέλα "τοποθετείται" στην κεντρική αύλακα, και για να λειτουργήσει απαιτεί διμερισμό του μορίου.

Ι.ΙΙ Πρωτεΐνες που δεσμεύουν Zn

Οι περιοχές (*domains*) αυτές διπλώνουν ανεξάρτητα από την υπόλοιπη πρωτεΐνη και απαιτούν οπωσδήποτε ψευδάργυρο, Zn^{+2} (*Lee et al.*, 1989) για να αλληλεπιδράσουν με το DNA. Ονομάζονται και δάκτυλοι ψευδαργύρου (*zinc fingers*) λόγω του τρόπου που θα μπορούσε να τα παραστήσει κάποιος στο χαρτί (*Freemont et al.*, 1991), δηλαδή με το ιόν του Zn^{+2} στη μέση να συνδέει αμινοξέα που βρίσκονται μακριά και η όλη αλληλουχία να παριστάνεται σαν μια θηλιά (Εικ 3). Για πρώτη φορά ανακαλύφθηκε μια τέτοια πρωτεΐνη, το 1985 σε βάτραχο (X*enopus laevis*), χαρακτηριζόμενη ως μεταγραφικός παράγοντας *IIIA (TFIIA) (Rhodes and Klug et al.,* 1986*).* Τα τελευταία χρόνια έχει λυθεί πειραματικά η δομή μερικών τέτοιων περιοχών με NMR σε διάλυμα (*Lee et al.*, 1989), και με κρυσταλλογραφία ακτίνων-Χ επίσης (*Pavletich and Pabo*, 1991). Όλες δείχνουν παρόμοια δομικά χαρακτηριστικά αποτελούμενα από μια αντιπαράλληλη β-πτυχωτή επιφάνεια (*β-sheet*), μια έλικα, και μια β-στροφή (*β-turn*) να τα συνδέει (*Freemont et al.*, 1991). Επειδή η παρούσα εργασία αφορά μελέτη τέτοιου είδους μορίων θα περιγράψουμε εκτενέστερα αυτή την οικογένεια.

Η δομή που έχει επιλυθεί πειραματικά με την βοήθεια πυρηνικού μαγνητικού συντονισμού (NMR) (*Lee et al.*, 1989) είναι η αλληλουχία ενός συνθετικού δακτύλου ψευδαργύρου και αντιστοιχούσε στον τριακοστό πρώτο δάκτυλο από την πρωτεΐνη Xfin σε βάτραχο (*Xenopus*) (*Altaba et al.*, 1987). Συγκεκριμένα επιλύθηκε η παρακάτω αλληλουχία:

YK**C**GL^{5}**C**ERSF^{10}VEKSA^{15}LSR**H**Q^{20}RV**H**KN

αφού κρίθηκε πιο αντιπροσωπευτική από ένα αριθμό 148 περίπου ομολόγων της (*Gibson et al.*, 1988). Πειράματα έδειξαν ότι ο παραπάνω δάκτυλος δεν δένεται επιλεκτικά σε κάποια αλληλουχία ενώ απουσία του ψευδαργύρου δεν δένεται καθόλου σε DNA. Με την επίλυση της παραπάνω δομής επιβεβαιώθηκαν οι προβλέψεις οι προβλέψεις για τη δευτεροταγή αλληλουχία των δακτύλων ψευδαργύρου (*Gibson et al.*, 1988).

Τα συμπεράσματα που προέκυψαν ήταν ότι ο ψευδάργυρος δένεται σε τετραεδρική διάταξη από δύο κυστεΐνες και δύο ιστιδίνες και συγκεκριμένα από δύο άτομα θείου και αζώτου αντίστοιχα. Η δομή αποτελείται από μια αντιπαράλληλη β-πτυχωτή επιφάνεια (*β-strand*), που περιέχει τις δύο κυστεΐνες που συγκρατούν τον ψευδάργυρο. Επίσης από μια α-έλικα που αρχίζει τρία περίπου αμινοξέα μετά την β-πτυχωτή επιφάνεια, και περιλαμβάνει και τις δύο ιστιδίνες που δένονται με τον ψευδάργυρο. Η έλικα αυτή παρουσιάζει μια 3_{10} διαμόρφωση περίπου στο τέλος της. Οι πλευρικές αλυσίδες που τοποθετούνται στο εσωτερικό του δακτύλου (π.χ. F^{10}, L^{16}, και

A^{15}) δημιουργούν ένα υδρόφοβο περιβάλλον για το ιόν του ψευδαργύρου.

Περιγράφοντας το παραπάνω μοντέλο θα μπορούσε να πει κανείς ότι μοιάζει με μια μικρή σφαιρική (*miniglobular*) πρωτεΐνη με καθορισμένο υδρόφοβο εσωτερικό, και τις πολικές και φορτισμένες πλευρικές αλυσίδες (π.χ. K^{13}, R^{18}, R^{21}) να δείχνουν στο εξωτερικό μέρος του μορίου. Για τα παραπάνω αμινοξέα της έλικας υπήρχαν υποψίες ότι μάλλον αλληλεπιδρούν με το σκελετό του DNA. Επίσης οι πλευρικές αλυσίδες E^{12}, S^{14}, S^{17}, E^{20} της έλικας προτάθηκαν ότι θα μπορούσαν να αλληλεπιδράσουν με τις βάσεις του DNA.

Συγκρίνοντας τη λυμένη δομή του δακτύλου ψευδαργύρου με τα μοντέλα που ήδη είχαν προταθεί (*Gibson et al.*, 1988 - *Berg et al.*, 1988) θα λέγαμε ότι μοιάζει αρκετά με τα προβλεπόμενα. Συγκεκριμένα το μοντέλο του Berg (*Berg et al.*, 1988) πλησιάζει πιο πολύ τη δομή που λύθηκε με NMR, διαφέρει όμως στο μήκος της έλικας (*Lee et al.*, 1989). Στο μοντέλο που προτάθηκε από τον Gibson (*Gibson et al.*, 1988) η θέση των β-κλώνων είναι λίγο μετατοπισμένη κατά δύο αμινοξέα, αλλά η θέση της έλικας προβλέπεται σωστά. Σε κανένα όμως από τα μοντέλα δεν είχε προβλεφθεί η ύπαρξη της 3_{10} διαμόρφωσης στο τέλος της έλικας.

Η αλληλουχία αυτού του τύπου των μορίων μετά από στατιστική ανάλυση (*Jacobs* 1992) μπορεί να χαρακτηριστεί γενικά ως:

C2-4X**C**11-12X**H**3-5X**H**

Το σύμπλοκο τριών δακτύλων ψευδαργύρου της πρωτεΐνης Zif 268 από ποντίκι με το αντίστοιχο DNA έχει λυθεί κρυσταλλογραφικά σε διακριτική ικανότητα 2.1 Angstroem (Pavlevitc and Pabo, 1991). Όπως φαίνεται στην Εικ.4 κάθε δάκτυλος αγκαλιάζει τη μεγάλη- αύλακα του DNA και χρησιμοποιεί το αμινοτελικό άκρο κάθε έλικας για να αλληλεπιδράσει ειδικά με τις αντίστοιχες βάσεις. Δεν υπάρχουν δεσμοί μεταξύ των δακτύλων και κάθε ένα από αυτά φαίνεται να είναι ανεξάρτητο από τα άλλα, αναγνωρίζοντας μια αλληλουχία 3 βάσεων αντίστοιχα. Η αναγνώριση της αλληλουχίας επιτυγχάνεται μέσω 11 δεσμών υδρογόνου μεταξύ των πλευρικών αλυσίδων των αμινοξέων της έλικας, και των αντιστοίχων βάσεων του DNA (Εικ.5). Αυτοί είναι κυρίως διπλοί υδρογονικοί δεσμοί μεταξύ της αργινίνης (*Arg*) και της γουανίνης (*G*) που πραγματοποιούνται μόνο με την μια αλυσίδα του DNA (Εικ.5).

Η μακριά πλευρική αλυσίδα της αργινίνης σταθεροποιείται από ένα ασπαρτικό οξύ (*Asp*) που βρίσκεται στη δεύτερη θέση της έλικας, με δύο χαρακτηριστικούς δεσμούς υδρογόνου (Εικ. 6). Κάτι επίσης αξιοσημείωτο είναι ότι μια από τις ιστιδίνες, με την οποία δένεται ο Zn^{+2}, αλληλεπιδρά με τον σκελετό του DNA. Συγκεκριμένα με το άτομο του αζώτου N^{δ}, ενώ το άζωτο N^{ε} δένεται με το ιόν του Zn^{+2}. Καταρχήν αφού ο ψευδάργυρος κατευθύνει την πλευρική αλυσίδα της ιστιδίνης και αυτή συνδέεται και με τον σκελετό του DNA, αυτό συνεπάγεται την έμμεση ανάμιξη του ψευδαργύρου στην αλληλεπίδραση. Άλλοι τέτοιοι δεσμοί πραγματοποιούνται από μια αργινίνη, στην β-πτυχωτή επιφάνεια και από μερικά άλλα αμινοξέα, μέσα σε έναν δάκτυλο που δρούν παρόμοια. Στο συγκεκριμένο παράδειγμα, λοιπόν, η αναγνώριση της συγκεκριμένης

αλληλουχίας του DNA συμβαίνει κυρίως από δεσμούς υδρογόνου μεταξύ αργινίνης - γουανίνης.

Εκτός από την παραπάνω οικογένεια υπάρουν και οι πρωτεΐνες της τάξης 2 (Πιν. 1), όπου οι υποκαταστάτες του Zn^{+2} είναι και οι τέσσερεις άτομα θείου από 4 κυστεΐνες. Αυτές οι πρωτεΐνες αποτελούνται κυρίως από τρεις περιοχές, δένονται στο DNA ως διμερή (Εικ. 7) και συνήθως ελέγχονται αλλοστερικά από στεροειδείς ορμόνες (Εικ. 8) (*Harrison*, 1991). Με NMR έχουν επιλυθεί πειραματικά η δομή του υποδοχέα του οιστρογόνου και του γλυκοκορτικοειδούς, το οποίο έχει επιλυθεί επίσης και κρυσταλλογραφικά (*Hard et al.*, 1990 - *Schwabe et al.*, 1990). Το δομικό πρότυπο συνθέτουν δύο έλικες που πακετάρονται σχεδόν κάθετα μεταξύ τους, και αναγνωρίζουν συμμετρικές αλληλουχίες DNA (*Harisson* 1991).

Συμπερασματικά θα λέγαμε ότι έχουν περιγραφεί αρκετά μοτίβα αλληλεπίδρασης πρωτεϊνών και DNA. Ίσως υπάρχουν αρκετά ακόμα και είναι πράγματι αξιοθαύμαστο πως η φύση έχει σχεδιάσει ένα τόσο πολύ ευαίσθητο μηχανισμό γονιδιακής ρύθμισης. Στη συνέχεια αυτής της εργασίας θα παρουσιάσουμε τα αποτελέσματα της στατιστικής μελέτης στις περιοχές που χαρακτηρίζονται ως δάκτυλοι ψευδαργύρου, και επίσης το μοντέλο του συμπλόκου της πρωτεΐνης CF2 με την αντίστοιχη αλληλουχία-στόχο DNA.

I.III Βιολογική λειτουργία της πρωτεΐνης CF2

Η πρωτεΐνη CF2 (chorion transcription factor) έχει χαρακτηριστεί ως μεταγραφικός παράγοντας γονιδίων στο χώριον (chorion) της δροσόφιλα. Η πρωτεΐνη εντοπίζεται στον πυρήνα (nuclei) των κυττάρων που σχηματίζουν το χώριον (*Hsu et al., unpublished data*). Δεν ανιχνεύεται στα πρώιμα εμβρυονικά στάδια (early embryonic stage) αλλά στα μετέπειτα στάδια της εμβρυωνικής ανάπτυξης (*later embryonic stage*). Συγκεκριμένα εντοπίζεται στις περιοχές που αντιστοιχούν στο μετέπειτα θώρακα (*thorax*) και στα γεννητικά όργανα (*abdominal*) (*Hsu et al., unpublished data*). Η πρωτεΐνη CF2 λειτουργεί ως διακόπτης ρυθμίζοντας την έκφραση γονιδίων στο χώριον κατά τη διάρκεια της ανάπτυξης (*development*) της δροσόφιλα.

Πρωτεΐνες τέτοιου είδους ρυθμίζονται με διάφορους τρόπους όπως i) μεταγραφικά ii) με μεταμεταγραφικούς παράγοντες (*posttranscriptional factors*), π.χ. εναλλακτική ωρίμανση (*alternative splicing*) iii) ή με παρουσία περιοριστικών παραγόντων (inhibitors) κ.λ.π. (*Hsu et al., 1992*). Στην περίπτωση της πρωτεΐνης CF2 που περιέχει περιοχές δακτύλων ψευδαργύρου συμβαίνει κάτι ανάλογο. (*Hsu et al., 1992*).

Κοντά στο αμινοτελικό άκρο της πρωτεΐνης βρίσκονται οι 2 πρώτοι δάκτυλοι (περιοχές 1 και 2) ενώ προς το καρβοξυτελικό άκρο της βρίσκονται οι 4 τελευταίοι (περιοχές 3, 4, 5 και 6). Ο ρόλος των δύο πρώτων δεν είναι ξεκάθαρος αφού παρουσία ή απουσία τους δεν παρατηρείται καμία διαφορά στην αλληλουχία DNA που αναγνωρίζεται συνολικά από την πρωτεΐνη (*Gogos et al., 1992*). Επίσης ο τρίτος δάκτυλος δεν παρουσιάζει συγγένεια για καμία συγκεκριμένη αλληλουχία DNA (*Gogos et al., 1992*). Μπορεί αυτοί οι δάκτυλοι απλώς να σταθεροποιούν την αλληλεπίδραση που επιτυγχάνεται από τους επόμενους τρεις (*Hsu et al., 1992*).

Μεταξύ της περιοχής του DNA που κωδικοποιεί για τον δάκτυλο 4 και 5 υπάρχει, και, μια άλλη κωδικοποιός περιοχή για δάκτυλο ψευδαργύρου η 5' η οποία δεν εμφανίζεται πάντα στην πρωτεΐνη. Αυτό οφείλεται στην ύπαρξη ιντρονίων (*introns*)

μεταξύ των περιοχών 4 και 5' και 5 που, δεν κωδικοποιούν για πρωτείνη παραλείπονται κατά την διεργασία της μεταγραφής από DNA σε RNA. Ανάλογα τώρα με το ποιες από αυτές τις περιοχές θα παραληφθούν[1] είναι δυνατόν η πρωτεΐνη να περιέχει ή όχι τον δάκτυλο 5' (*Hsu eta al., 1992*). ο δάκτυλος αυτός έχει στις κρίσιμες θέσεις (αυτές δηλαδή που αλληλεπιδρούν με τις βάσεις του DNA) ακριβώς τα ίδια αμινοξέα με το δάκτυλο 5 (*Hsu eta al.*, 1992). Παρουσία και του δακτύλου 5' η αλληλουχία ΟΝΑ που επιλέγεται από την πρωτεΐνη CF2 αλλάζει. (*Hsu eta al., 1992*). Η πρωτεΐνη ανιχνεύεται και στις δύο μορφές (με ή χωρίς τον δάκτυλο 5') σε διαφορετικές ποσότητες στα εμβρυϊκά στάδια ανάπτυξης της δροσόφιλα, έχοντας την δυνατότητα με αυτόν τον τρόπο να ρυθμίζει την έκφραση διαφορετικών γονιδίων στο χώριον (*Hsu eta al., 1992*).

Η αλληλουχία DNA που αναγνωρίζεται από την πρωτεΐνη είναι η ακόλουθη:

5'-GTATATATA-3'

Συγκεκριμένα οι τριάδες GTA, TAT ATA αναγνωρίζονται αντίστοιχα από τους δακτύλους 6, 5 και 4 αντίστοιχα ο δάκτυλος 5' αναγνωρίζει την ίδια αλληλουχία με τον 5 και έτσι η αλληλουχία DNA που αναγνωρίζεται όταν υπάρχει και αυτός ο δάκτυλος είναι η 5'-GTATATTATATA-3' (*Gogos et al., 1992*).

I.IV Λόγοι πραγματοποίησης αυτής της εργασίας

Η κρισιμότητα και το βιολογικό ενδιαφέρον των πρωτεϊνών που αλληλεπιδρούν με το DNA είναι τεράστιο αφού ρυθμίζουν βασικές διεργασίες, όπως η ανάπτυξη η διαφοροποίηση και η διαίρεση του κυττάρου. Η κατανόηση λοιπόν του μηχανισμού δράσης τους σε μοριακό επίπεδο θα δώσει απαντήσεις σε καίρια βιολογικά ερωτήματα. Μέχρι τώρα έχει λυθεί κρυσταλλογραφικά η δομή τριών δακτύλων ψευδαργύρου με το αντίστοιχο DNA (*Pavletich and Pabo, 1991*) και δύο δακτύλων ψευδαργύρου με NMR (*Lee et al.*, 1989 - *Klevit et al., 1990*). Παράλληλα έχουν γίνει πειράματα για την εύρεση του κώδικα αναγνώρισης μεταξύ πρωτεΐνης και του DNA (*Desjarlais and Berg, 1992 - Nardelli et al., 1992 - Jacobs1992*). Προς το παρόν είναι γνωστοί μόνο κάποιοι κανόνες συγγένειας μερικών αμινοξέων με συγκεκριμένες βάσεις του DNA, όχι όμως και ο πλήρης κώδικας αναγνώρισης. Επίσης έχουν γίνει και κάποιες στατιστικές μελέτες σε

[1] *Κατά την διεργασία της ωρίμανσης (splicing) του DNA παραλείπονται οι περιοχές που δεν κωδικοποιούν για πρωτεΐνη, δηλαδή τα ιντρόνια. Είναι δυνατόν όμως κατά την διάρκεια της παραπάνω διεργασίας να παραληφθούν και περιοχές του DNA που κωδικοποιούν για πρωτεΐνη και βρίσκονται μεταξύ των ιντρονίων. Κάτι ανάλογο συμβαίνει και στην περίπτωση που απουσιάζει ο δάκτυλος 5' από την πρωτεΐνη.*

αλληλουχίες δακτύλων ψευδαργύρου (*Jacobs1992 - Desjarlais and Berg, 1992*).

Ο Jacobs (*Jacobs, 1992*) μελέτησε αυτές τις αλληλουχίες ερευνώντας πόσο συντηρημένες ή μη είναι όλες οι θέσεις ενός δακτύλου ψευδαργύρου σε ομάδες πρωτεϊνών που απέχουν ή όχι εξελικτικά. Συμπερασματικά κατέληξε ότι τα αποτελέσματα διαφέρουν σε διαφορετικές οικογένειες πρωτεϊνών. Πιο συγκεκριμένα τα μόνα αμινοξέα που συντηρούνται σε όλες τις πρωτεΐνες είναι οι δύο κυστεϊνες και οι δύο ιστιδίνες που δένονται με τον ψευδάργυρο. Οι μη συντηρημένες θέσεις που αλληλεπιδρούν ειδικά με τις βάσεις του DNA (*Pavletich and Pabo, 1991*) συντηρούνται σε σχετικά υψηλά ποσοστά σε ομόλογες πρωτεΐνες. Αυτό συμβαίνει γιατί ίσως οι αλληλουχίες των δακτύλων ψευδαργύρου έχουν προέλθει από αντιγραφή κάποιων γονιδίων (*gene duplication*) μέσα στους διάφορους οργανισμούς. Αυτό εξηγεί το γεγονός γιατί αυτές οι θέσεις συντηρούνται. σχετικά σε ομόλογες πρωτεΐνες και μεταβάλλονται αρκετά σε μη ομόλογες.

Τα αμινοξέα του συνδέσμου (*linker*) συντηρούνται επίσης σε υψηλά ποσοστά στις ομόλογες ή μη πρωτεΐνες. Τέλος οι θέσεις που τοποθετούν τις πλευρικές τους αλυσίδες προς το εξωτερικό των δακτύλων έχουν ίδια ποσοστά (σχετικά χαμηλά) σε όλες τις οικογένειες πρωτεϊνών.

Σε αυτή την ανάλυση δεν μελετήθηκε όμως ποια συγκεκριμένα αμινοξέα, ή συνδυασμοί τους απαντούν στις κρίσιμες (αυτές δηλαδή που αλληλεπιδρούν ειδικά με το DNA) ή μη θέσεις της αλληλουχίας ενός δακτύλου ψευδαργύρου.

Επίσης η ομάδα *Desjarlais* και *Berg* (*Desjarlais and Berg, 1992*) εξέτασε μια άλλη πλευρά του θέματος. Συγκεκριμένα μελετήθηκε στατιστικά μια βάση δεδομένων από δακτύλους ψευδαργύρου και αναφέρονται τα αποτελέσματα για δύο μόνο θέσεις από την αλληλουχία τους. Πρέπει να σημειωθεί ότι αναφέρονται απλώς πόσες φορές παρατηρούνται συγκεκριμένα αμινοξέα στις δύο αυτές θέσεις χωρίς περαιτέρω στατιστική ανάλυση. Επίσης έγιναν και κάποια πειράματα με την αλληλουχία τριών δακτύλων ψευδαργύρου της πρωτεΐνης *Sp1* από άνθρωπο σε προσπάθεια ανεύρεσης του κώδικα αναγνώρισης του DNA.

Τα αποτελέσματα ανέδειξαν κάποιους μερικούς γενικούς κανόνες συγγένειας κάποιων αμινοξέων με βάσεις του DNA. Σε καμία περίπτωση όμως δεν μπορεί να προβλεφθεί με ακρίβεια η αλληλουχία του DNA που μπορεί να αναγνωριστεί από μια συγκεκριμένη πρωτεΐνη ή και το αντίθετο.

Προς το παρόν υπάρχουν πολλά ακόμα ερωτηματικά πάνω στο αντικείμενο, όπως η ύπαρξη κάποιου προτύπου στο τρόπο αλληλεπίδρασης των δακτύλων ψευδαργύρου με το DNA, ή η συσχέτιση της πρωτεϊνικής ακολουθίας και άλλων δεδομένων, όπως για παράδειγμα τα πόσο συντηρημένη είναι κάθε θέση, με την

βιολογική λειτουργία τους. Φυσικά παραμένει και το ερώτημα του κώδικα αναγνώρισης μεταξύ των δακτύλων ψευδαργύρου και του DNA. Όλα τα παραπάνω βοηθούν στην καλύτερη κατανόηση πολύπλοκων βιολογικών φαινομένων όπως η γονιδιακή ρύθμιση, κ.λ.π. Από εκεί και πέρα θα μπορούσαν να εξυπηρετήσουν και άλλους σκοπούς, όπως η αντιμετώπιση γενετικών ασθενειών ή γενικότερα άλλων ανωμαλιών που οφείλονται στην ελαττωματική λειτουργία τέτοιων πρωτεϊνών.

ΚΕΦΑΛΑΙΟ ΙΙ

ΠΕΙΡΑΜΑΤΙΚΑ

II. Μοντέλο του CF2 με το DNA και στατιστική ανάλυση αλληλουχιών δακτύλων Zn

II.I Γενικά

Στην εισαγωγή αναφέρθηκε, με λεπτομέρεια, το μοντέλο αλληλεπίδρασης της πρωτεΐνης Zif268 (*Pavletitc and Pabo,* 1991) με το αντίστοιχο DNA, που ανήκει στην κατηγορία CC/HH. Για κάθε δάκτυλο, αμινοξέα από την α-έλικα αλληλεπιδρούν ειδικά με το DNA και η αλληλουχία της πρωτεΐνης τοποθετείται αντιπαράλληλα με την αλληλουχία του DNA. Ο λόγος που χρησιμοποιήσαμε το παραπάνω μοντέλο είναι γιατί αποτελεί χαρακτηριστικό παράδειγμα τέτοιου είδους αλληλεπιδράσεων (*Pavletitc and Pabo,* 1991). Το ενδιαφέρον είναι μεγάλο για να δούμε αν πράγματι διατηρούνται τα γενικά και τα ειδικά χαρακτηριστικά της αλληλεπίδρασης αφού στην προκειμένη περίπτωση το DNA στόχος αποτελείται κυρίως από αλληλουχία βάσεων A-T, και όχι από G-C όπως στο μοντέλο του Zif268/DNA.

Στην στατιστική ανάλυση αλληλουχιών δακτύλων ψευδαργύρου προσπαθήσαμε να δούμε ποιές θέσεις είναι συντηρημένες ή μη, αν υπάρχουν προτιμήσεις αμινοξέων σε κρίσιμες θέσεις που αλληλεπιδρούν ειδικά με το DNA, κ.λ.π. Γενικά προσπαθήσαμε να καταλάβουμε ποιός είναι ο συγκεκριμένος ρόλος κάθε αμινοξέος στην αλληλουχία ενός δακτύλου, και αν αυτό ισχύει για καθένα από αυτά χρησιμοποιώντας παράλληλα και ως οδηγό το μοντέλο του CF2

II.II ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

II.II.1 Μοντέλο CF2 με το DNA

Η δημιουργία μοντέλου της τρισδιάστατης δομής μιας πρωτεΐνης βασίζεται σε μία ή περισσότερες λυμένες δομές ομολόγων πρωτεϊνών. Χαρακτηρίζουμε έτσι πρωτεΐνες που παρουσιάζουν ομολογία τόσο στην αλληλουχία τους, όσο και στην βιολογική τους λειτουργία. Η τεταρτοταγής δομή σε μια οικογένεια ομολόγων πρωτεϊνών είναι γενικά η ίδια (*Overington et al., 1990*).Υπάρχουν όμως περιοχές που η δομή τους μεταβάλλεται (*structurally variable regions*) και μερικές που η δομή τους συντηρείαι (*structurally conserved regions*) στα διαφορετικά μέλη της οικογένειας (*Overington et al., 1990*). Όταν η ομολογία με τη λυμένη δομή της πρωτεΐνης δεν είναι αρκετά μεγάλη απαιτείται επιπλέον προσοχή ώστε τα αποτελέσματα να μην απέχουν

πολύ από την πραγματικότητα (*Novotny et al., 1984*).

Το τρισδιάστατο μοντέλο των δακτύλων 4, 5 και 6 του CF2 βασίστηκε πάνω στις συντεταγμένες του Zif268 με το DNA, που μας παραχωρήθηκαν ευγενικά από τον κ.κ. *C.Pabo*, και πραγματοποιήθηκε με την βοήθεια των υπολογιστών μοριακών γραφημάτων *E&S PS390* και *ESV10.* Τα προγράμματα που χρησιμοποιήσαμε ήταν το *FRODO* (*Jones*, *1988*), το *O* (*Jones*, *1991*), και το *SYBYL* (TRIPOS ASSOCIATES, Inc.)[2] για την περαιτέρω ανάλυση της δομής του μοντέλου. Η μοντελοποίηση για κάθε έναν από τους δακτύλους (*finger*) 4,5, και 6, του CF2 έγινε βάσει των δακτύλων 1,2,3 του Zif268 αντίστοιχα. Οι διατάξεις των συνδέσμων (*linkers*) μεταξύ των δακτύλων διατηρήθηκαν σταθερές, όπως στο μοντέλο του Zif268. Η αναγνώριση συγκεκριμένων θέσεων πάνω στην αλληλουχία του CF2 έγινε βάσει των ήδη γνωστών για αυτήν την οικογένεια πρωτεϊνών (*Desjarlais and Berg, 1992, Jacobs, 1992*).

Η μοντελοποίηση έγινε κρατώντας σταθερό τον σκελετό των πρωτεϊνών και αντικαθιστώντας συγκεκριμένα αμινοξέα στις αντίστοιχες θέσεις. Μια επιπλέον αλλαγή χρειάστηκε στην περίπτωση του δακτύλου 4 γιατί ο αντίστοιχος δάκτυλος του Zif268 περιείχε δύο παραπάνω αμινοξέα, μεταξύ του βρόγχου (*loop*) των δύο κυστεϊνών που συνδένονται με το ιόν του Zn^{+2}. Για αυτή την περιοχή χρησιμοποιήθηκε η αντίστοιχη διαμόρφωση των δακτύλων 2, ή 3, που έχουν τον ίδιο αριθμό αμινοξέων με το δάκτυλο 4 και μοιάζουν πολύ μεταξύ τους (η μέση διαφορά των συντεταγμένων των ατόμων τους, *rms deviation* ήταν 0.134 A). Λόγω του γεωμετρικού περιορισμού του πολύ μικρού μήκους τους οι συγκεκριμένες στροφές θα μπορούσαν ίσως να αποτελέσουν αντιπροσωπευτικό παράδειγμα για την σύνδεση της β πτυχωτής επιφάνειας (*β-sheet*), των δακτύλων ψευδαργύρου. Επίσης στο δάκτυλο 6 υπάρχει ένα επιπλέον αμινοξύ στην έλικα μεταξύ των δύο ιστιδινών που συνδέονται με τον Zn^{+2}, σε σχέση με το δάκτυλο 3 του Zif268. Για να αντιμετωπιστεί αυτή η διαφορά και για να διατηρηθεί η σωστή γεωμετρία του Zn^{+2}, δώσαμε ένα χαρακτήρα περισσότερο α-ελικοειδή στην 3_{10} διαμόρφωση που ήδη υπήρχε. Η μέση διαφορά των συντεταγμένων όλων των ατόμων αυτής της περιοχής, πριν και μετά την αλλαγή ήταν 0.234 A. Χρησιμοποιώντας την ρουτίνα του FRODO για γεωμετρικές κανονικοποιήσεις διατηρήσαμε την γεωμετρία σε αποδεκτά επίπεδα[3] , κατά την διάρκεια όλων αυτών των αλλαγών.

Με την βοήθεια του προγράμματος SYBYL κατασκευάσαμε θεωρητικά την αλληλουχία του DNA στόχου σε B διαμόρφωση. Αφού βρήκαμε το σωστό προσανατολισμό του σε σχέση με την πρωτεΐνη, χρησιμοποιώντας την ρουτίνα των ελαχίστων τετραγώνων (*lsq procedure*) του προγράμματος O (*Jones*, *1991*)

[2] (1699 S. Hardley Rd., Suite 303 St.Louis, Missouri 63144-2913)

[3] *σύμφωνα δηλαδή με τα στερεοχημικά δεδομένα που ισχύουν για τις πρωτεΐνες*

προχωρήσαμε στην κατασκευή μοντέλου του CF2 με το DNA. Συγκεκριμένα αυτό έγινε στοιχίζοντας τα αντίστοιχα άτομα (όσο το δυνατόν καλύτερα) των δύο ολιγονουκλεοτιδίων. Ανάλογη διαδικασία ακολουθήσαμε επίσης και για να συγκρίνουμε το μοντέλο του DNA, που δημιουργήσαμε με το *SYBYL* με μία πειραματικά προσδιορισμένη δομή B-DNA με αλληλουχία CGACATATATTGCG (*Yoon et al., 1988*). Τα δύο μόρια έχουν κοινό το εξαμερές (ATATAT) και η σύγκριση έγινε για να ελέγξουμε την ποιότητα θεωρητικού DNA. Μετά από τις απαραίτητες διορθώσεις πραγματοποιήσαμε ελαχιστοποίηση της ενέργειας (*energy minimazation procedure*) του συμπλόκου χρησιμοποιώντας την αντίστοιχη ρουτίνα του προγράμματος *SYBYL*. Ας σημειωθεί εδώ ότι το φορτίο που χρησιμοποιήθηκε και για τους τρείς Zn ήταν +2, όπως και στην πραγματικότητα. Μετά από πολλούς κύκλους το μοντέλο σταθεροποιήθηκε σε ένα ελάχιστο.

Η βελτίωση της τεταρτοταγούς δομής μιας πρωτεΐνης είναι δυνατόν να πραγματοποιηθεί από διεργασίες "ελαχιστοποίησης της ενέργειας". Οι δυνάμεις μεταξύ των ατόμων είναι δυνατόν να περιγραφούν από ημιεμπειρικές συναρτήσεις κατάλληλα προσαρμοσμένες για την περίπτωση των μακρομορίων. Είναι λοιπόν δυνατόν να προσεγγίσουμε τη πραγματική δομή μιας πρωτεΐνης μετακινώντας τα άτομα που την αποτελούν με τέτοιο τρόπο, ώστε η τιμή της συνάρτησης της δυναμικής της ενέργειας να ελαχιστοποιείται (*Lifson, 1973, 1980*). Η ενέργεια Ε μπορεί να εκφραστεί ως άθροισμα από διάφορους όρους ως εξής:

Οι μεταβλητές εΙναι τα μήκη δεσμών b, οι γωνΙες δεσμών θ, οι δίεδρες γωνίες φ, και οι αποστάσεις μεταξύ των ατόμων r. Όλα τα παραπάνω εξαρτώνται από τις ατομικές συντεταγμένες χ,y,z, που ορίζονται για μια συγκεκριμένη διάταξη της πρωτεΐνης σ' ένα σύστημα καρτεσιανών συντεταγμένων. Οι διαφορές παράμετροι kb, bo, kθ, θo, kφ, η, δ, ε, r, q καθορίζουν ποσοτικά την συνάρτηση της ενέργειας. Οι τιμές τους καθορίζονται όταν χρησιμοποιήσουμε την παραπάνω συνάρτηση για υπολογισμό σταθερών κυψελΙδας (*unit cells*) σε κρυστάλλους, τιμές ισορροπίας για μήκη δεσμών ή γωνιών, κ.λ.π. Στην συνέχεια προσαρμόζουμε αυτές τις παραμέτρους για να

προσεγγίσουμε ανάλογα τα πειραματικά δεδομένα (*Levitt and Lifson, 1969*).

Η παραπάνω μέθοδος ελαχιστοποίησης της ενέργειας παρουσιάζει κάποια μειονεκτήματα όπως: i) η πρωτεΐνη όταν έχει την πραγματική της δομή (*native form*) μπορεί να μην έχει την διάταξη που αντιστοιχεί στην ελάχιστη τιμή της δυναμικής της ενέργειας ii) οι διαφορές στην τιμή της ενέργειας που ανιχνεύονται κατά την διεργασία της ελαχιστοποίησης είναι πολύ μικρές σε σχέση με τη τάξη μεγέθους αυτής, κάτι που δυσκολεύει εξαιρετικά αυτήν την διεργασία iii) Η μορφή της παραπάνω συνάρτησης παριστάνεται από μια υπερεπιφάνεια με μεγάλο αριθμό τοπικών ελαχίστων, που οφείλεται στην μορφή της. Το γεγονός αυτό δυσκολεύει, εξαιρετικά την εύρεση του ολικού ελαχίστου (*global minimum*)[4] .

ΙΙ.ΙΙ.2 Συλλογή και στατιστική ανάλυση των αλληλουχιών δακτύλων ψευδαργύρου

Στατιστική ανάλυση σαν αυτή που πραγματοποιήσαμε, απαιτεί όσο το δυνατόν μεγαλύτερη συλλογή αλληλουχιών δακτύλων ψευδαργύρου. Για αυτό τον σκοπό αναπτύξαμε ένα πρόγραμμα που δημιουργεί μια βάση δεδομένων από απομονωμένες τέτοιες περιοχές. Το πρόγραμμα λειτουργεί ως εξής:

Πρώτα από όλα απομονώνει όλες τις πρωτεΐνες που έχουν αναφερθεί στην βάση δεδομένων, *Swiss database* ως "zinc finger domain proteins", πρωτεΐνες δηλαδή που περιέχουν περιοχές με δακτύλους ψευδαργύρου. Από αυτές τις πρωτεΐνες απομονώσαμε ακριβώς όσες περιοχές είχαν την αλληλουχία:

C2-4X**C**11-13X**H**3-5X**H**

αφού σύμφωνα κατά τα γνωστά (*Harisson, 1991*) αυτή είναι η χαρακτηριστική τέτοιων περιοχών. Η πλειοψηφία τους 577, είχαν την αλληλουχία της μορφής:

C2X**C**12X**H**3X**H**

με μια μόνο σημαντική εξαίρεση, όπου βρέθηκαν εξήντα δάκτυλοι ψευδαργύρου της μορφής:

C2X**C**12X**H**4X**H**

[4] για περισσότερα δες "Folding and association of proteins" *(R. JAENICKE)*

και κάποιες άλλες πολύ μικρές εξαιρέσεις. Οι πρωτεΐνες από όπου ανήκουν οι αλληλουχίες προέρχονται κυρίως από άνθρωπο, ποντικό, και βάτραχο, και δευτερευόντως από σκουλίκι, μύκητα, κοτόπουλο, κ.λ.π. (δες Πιν. 2 στο Παράρτημα)

II.II.3 Μεθοδολογία

Μετά από την αντιστοίχηση των επτά θέσεων **13-19** στην έλικα που έγινε βάσει των ήδη γνωστών για αυτήν την οικογένεια πρωτεϊνών[5] (*Desjarlais and Berg, 1992, Jacobs 1992*) υπολογίσαμε τις συχνότητες για κάθε θέση σε όλους τους δάκτυλους ψευδαργύρου της βάσης δεδομένων. Αυτό έγινε χρησιμοποιώντας προγράμματα που αναπτύχθηκαν από μας. Ο λόγος που υπολογίζουμε αυτά τα μεγέθη είναι για να αποδείξουμε στατιστικά ποιά ζευγάρια, ή τριάδες, αμινοξέων συνδυάζονται κατά τύχη, και ποιά όχι.

Ορίζουμε ως **R**(r,p) τον αριθμό που παρατηρείται το συγκεκριμένο αμινοξύ **r** στην θέση **p** στις αλληλουχίες των δακτύλων ψευδαργύρου. Επίσης ορίσαμε τις συχνοτήτες (*occurence*) **O**(r,p) όπως φαίνεται παρακάτω:

$$\mathbf{O}(r,p) = \frac{\mathbf{R}(r,p)}{\sum_{r=1}^{20} \mathbf{R}(r,p)} \qquad (1)$$

όπου **r** ο αριθμός όλων των αμινοξέων που απαντούν σε μια συγκεκριμένη θέση και **p** οποιαδήποτε θέση μέσα στην αλληλουχία ψευδαργύρου.

Για όλα τα δυνατά ζευγάρια, μεταξύ των θέσεων, **13-19**, υπολογίσαμε τις παραπάνω συχνότητες και τις τυπικές αποκλίσεις τους (*e.s.d's*) βάσει της:

$$\sigma(r,p) = \sqrt{\mathbf{R}(r,p)}^{6} \qquad (2)$$

Ορίζουμε ως $\mathbf{J}(r_1/p_1,r_2/p_2)$ τον αριθμό που παρατηρούνται ταυτόχρονα τα αμινοξέα $\mathbf{r_1}$, $\mathbf{r_2}$ στις θέσεις $\mathbf{p_1}$, $\mathbf{p_2}$ αντίστοιχα. Το κανονικοποιημένο μέγεθος ανάλογα θα δίνεται από την σχέση:

---

[5] *Υπενθυμίζουμε ότι αυτή η αρίθμηση συμπίπτει με την αρχή της έλικας και αντιστοιχεί με τις θέσεις s_3-m_3.*

[6] *όπως ορίζεται από τη θεωρία της στατιστικής των μετρήσεων (count statistics)*

$$\mathbf{J}'(r_1/p_1,r_2/p_2) = \frac{\mathbf{J}(r_1/p_1,r_2/p_2)}{\sum_{r_1,r_2} \mathbf{J}(r_1/p_1,r_2/p_2)} \qquad (3)$$

Το αντίστοιχο μέγεθος των σχετικών συχνοτήτων (*relative occurence*) $\mathbf{R.O.}(r_1/p_1,r_2/p_2)$ βάσει της (1) μπορεί να οριστεί ως εξής:

$$\mathbf{R.O.}(r_1/p_1,r_2/p_2) = \frac{\mathbf{J}'(r_1/p_1,r_2/p_2)}{\mathbf{O}(r_1,p_1)\cdot\mathbf{O}(r_2,p_2)} \qquad (4)$$

Δεδομένου ότι οι ποσότητες $\mathbf{R}(r,p)$ και $\mathbf{R.O.}(r_1/p_1, r_2/p_2)$ συνδέονται από τις (1)-(4), προκύπτει από την θεωρία σφαλμάτων ότι η απόκλιση του $\mathbf{R.O.}(r_1/p_1, r_2/p_2)$ δίνεται από την σχέση:

$$\sigma(r_1/p_1,r_2/p_2) = \mathbf{N}\cdot\frac{\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}}\cdot\sqrt{1/\mathbf{a}+1/\mathbf{b}+1/\mathbf{c}} \qquad (5)$$

$$\mathbf{N} = \sum_{r} \mathbf{R}(r,p) \qquad (6)$$

$$\mathbf{c}=\mathbf{J}(r_1/p_1,r_2/p_2) \qquad (7)$$

$$\mathbf{a} = \mathbf{R}(r_1/p_1) \qquad (8)$$

$$\mathbf{b} = \mathbf{R}(r_2/p_2) \qquad (9)$$

Για τις τριάδες μπορούμε να ορίσουμε ως $\mathbf{K}(r_1/p_1,r_2/p_2,r_3/p_3)$ τον αριθμό των μετρήσεων όταν τα αμινοξέα $\mathbf{r_1,r_2,r_3}$, παρατηρούνται ταυτόχρονα στις θέσεις $\mathbf{p_1,p_2,p_3}$. Ο κανονικοποιημένος πίνακας $\mathbf{K}'(r_1/p_1,r_2/p_2,r_3/p_3)$ θα δινεται από την παρακάτω σχέση, κατά αναλογία με την (2):

$$\mathbf{K}'(r_1/p_1,r_2/p_2,r_3/p_3) = \frac{\mathbf{K}(r_1/p_1,r_2/p_2,r_3/p_3)}{\sum_{r_1,r_2,r_3} \mathbf{K}(r_1/p_1,r_2/p_2,r_3/p_3)} \qquad (10)$$

και τις σχετικές συχνότητες $\mathbf{R.O.}(r_1/p_1,r_2/p_2,r_3/p_3)$ αντίστοιχα με την (4) τις ορίσαμε ως εξής:

$$\mathbf{R.O.}(r_1/p_1,r_2/p_2,r_3/p_3) = \frac{\mathbf{K'}(r_1/p_1,r_2/p_2,r_3/p_3)}{\mathbf{O}(r_1/p_1)\cdot\mathbf{O}(r_2,p_2)\cdot\mathbf{O}(r_3,p_3)} \qquad \mathbf{(11)}$$

επίσης οι τυπικές αποκλίσεις των παραπάνω ποσοτήτων ορίζονται αντίστοιχα ως:

$$\sigma(r_1/p_1,r_2/p_2,r_3/p_3) = \mathbf{N}^2\cdot\frac{\mathbf{d}}{\mathbf{a}\cdot\mathbf{b}\cdot\mathbf{c}}\cdot\sqrt{(1/\mathbf{a}+1/\mathbf{b}+1/\mathbf{c}+1/\mathbf{d})} \qquad \mathbf{(12)}$$

όπου το **N** ορίζεται από την (6) και

$$\mathbf{d} = \mathbf{K}(r_1/p_1,r_2/p_2,r_3/p_3) \qquad \mathbf{(13)}$$

$$\mathbf{a} = \mathbf{R}(r_1/p_1) \qquad \mathbf{(14)}$$

$$\mathbf{b} = \mathbf{R}(r_2/p_2) \qquad \mathbf{(15)}$$

$$\mathbf{c} = \mathbf{R}(r_3/p_3) \qquad \mathbf{(16)}$$

Επίσης ορίζουμε ως πολυμορφικότητα (*variance*) $\mathbf{var(p)}$ ενός αμινοξέος σε μια θέση **p** ως εξής:

$$\mathbf{var(p)} = 1 - \max \mathbf{O(p)} \qquad \mathbf{(17)}$$

Την παραπάνω ποσότητα την υπολογίσαμε για όλα τα αμινοξέα των δακτύλων στην βάση δεδομένων. Επιχειρούμε στην συνέχειανα χρησιμοποιώντας τις παραπάνω ποσότητες να βρούμε τους κανόνες, αν υπάρχουν που διέπουν την κατανομή των αμινοξέων σε μία αλληλουχία ενός δακτύλου ψευδαργύρου.

ΚΕΦΑΛΑΙΟ ΙΙΙ

ΑΠΟΤΕΛΕΣΜΑΤΑ

III. ΑΠΟΤΕΛΕΣΜΑΤΑ

III.I Μοντέλο του συμπλέγματος του CF2 με το DNA

Το τρισδιάστατο μοντέλο των τριών δακτύλων ψευδαργύρου του CF2, παρά τις διαφορές που παρουσιάζει με το αντίστοιχο του Zif268, βασίστηκε αποκλειστικά σ' αυτό. Εφαρμόζοντας τις διάφορες διορθώσεις, που αναφέρθηκαν παραπάνω η ανάπτυξη του μοντέλου ήταν άμεση από αυτό του Zif268. Το θεωρητικό κομμάτι του DNA, που δημιουργήσαμε με την βοήθεια του προγράμματος *SYBYL*, δεν απέχει πολύ από το πειραματικό κομμάτι των έξι βάσεων, που είναι κοινό και στα δύο. Συγκεκριμένα η μέση διαφορά (*rms deviation*) των συντεταγμένων όλων των ατόμων τους, είναι λιγότερη από 1.5 Angstroem. Η μεγαλύτερη διαφορά τους εντοπίζεται στο ότι οι βάσεις του πειραματικού μοντέλου δεν είναι απόλυτα συνεπίπεδες, (*propeller twist*)[7] , κάτι που δεν συμβαίνει για το θεωρητικό. Για τον σκοπό όμως μιας ποιοτικής περιγραφής των αλληλεπιδράσεων του CF2 με το DNA οι διαφορές των δύο μοντέλων είναι ελάχιστες.

Όλες οι αλληλεπιδράσεις που προβλέπονται βάσει του μοντέλου μας φαίνονται στην Εικ. 10. Επίσης η σχετική θέση του δακτύλου 5 ως προς το DNA φαίνεται στην Εικ. 11. Οι ειδικές αλληλεπιδράσεις περιλαμβάνουν δεσμούς υδρογόνου μεταξύ των πλευρικών αλυσίδων, που βρίσκονται στο αμινοτελικό άκρο της έλικας των δακτύλων και των βάσεων του DNA. Επίσης οι δεσμοί με τον σκελετό του DNA πραγματοποιούνται κατά αναλογία με το μοντέλο του Zif268, απλά οι λυσίνες έχουν αντικαταστήσει τις αντίστοιχες αργινίνες Εικ. 12. Οι αποστάσεις των ατόμων, δέκτη-δότη, που σχηματίζουν δεσμούς υδρογόνου με τις βάσεις του DNA, είναι οι κατάλληλες (μικρότερες από 3.2 Angstroem) μόνο για τον δάκτυλο 5. Για τον τέταρτο και τον έκτο όμως είναι μεγαλύτερες από 3.5 Angstroem, στο δομικό πλαίσιο που βασίζεται στο Zif268 αλλά οι σχετικοί προσανατολισμοί πρωτεΐνης και DNA είναι κατάλληλοι για τον σχηματισμό υδρογονικών δεσμών. Αυτό οφείλεται μάλλον στο ότι οι πλευρικές αλυσίδες που παίρνουν μέρος σ' αυτήν την διεργασία είναι μικρότερες από ότι στο μοντέλο του Zif268, όπου κυριαρχούσαν οι αργινίνες.

[7] *Principles of Nucleic Acid Structure, Springer Verlag*

Οι πλευρικές αλυσίδες του δακτύλου 5 (τυροσίνη, λυσίνη) είναι παρόμοιου μήκους με αυτές της αργινίνης γι' αυτό και οι αποστάσεις, για σχηματισμό υδρογονοδεσμών είναι οι κατάλληλες. Και για τους υπόλοιπους δακτύλους είναι δυνατόν να επιτευχθούν παρόμοια, αν μετακινηθεί ολόκληρος ο δάκτυλος προς το μέρος του DNA σε διεύθυνση κάθετη προς τον άξονα της α-έλικας. Μετακίνηση κατά περίπου 1 Angstroem είναι αρκετή για να διορθωθούν οι αποστάσεις των πλευρικών αλυσίδων από τις βάσεις του DNA. Αποτέλεσμα αυτής της μετακίνησης είναι η πιο "συμπαγής" τοποθέτηση των δακτύλων του CF2, μέσα στην κύρια αύλακα. Όλα τα εσωτερικά δομικά χαρακτηριστικά παραμένουν αμετάβλητα, όπως επίσης και οι υδρογονικοί δεσμοί προς το σκελετό του DNA.

Η μόνη δομική αλλαγή που απαιτείται είναι στην διαμόρφωση του συνδέσμου (*linker*) *GEK* που συνδέει τους δακτύλους. Χρησιμοποιώντας την ρουτίνα *DGNL*[8] του *FRODO* παρατηρήσαμε ότι ήταν δυνατόν να συνδεθούν οι δύο δάκτυλοι με διαμόρφωση συνδέσμου τέτοια, που να υπάρχει και σε λυμένες πρωτεϊνικές δομές. Είναι λοιπόν πιθανόν η δομή του συνδέσμου να παρουσιάζει σχετική ευελιξία (*flexibility*) ώστε κάθε δάκτυλος να μπορεί να τοποθετείται στην σωστή απόσταση σχετικά με το DNA.

Οι αλληλεπιδράσεις του μοντέλου CF2 με το DNA γενικά συμφωνούν με αυτές του Zif268. Η πλειοψηφία των ειδικών αλληλεπιδράσεων γίνεται με την αλυσίδα του DNA που τοποθετείται αντιπαράλληλα, προς την αλληλουχία της πρωτεΐνης. Λόγω του ότι το μοντέλο είναι θεωρητικό δεν μπορούμε να πούμε με ακρίβεια ποιές αλληλεπιδράσεις από την Εικ. 10 συμβαίνουν στην πραγματικότητα. Από τα πειράματα μεταλλαξογένεσης φαίνεται ότι για τον δάκτυλο 4 δεν λαμβάνουν χώρα όλες οι αλληλεπιδράσεις. Βάσει του μοντέλου Zif268/DNA πραγματοποιούνται δύο τουλάχιστον, ειδικές αλληλεπιδράσεις ανά δάκτυλο, και κάτι ανάλογο φαίνεται να συμβαίνει και στους δακτύλους 5 και 6 του CF2. Διαφορά παρουσιάζεται στην δυνατότητα αλληλεπίδρασης του CF2 με βάσεις από την αλυσίδα του DNA που τοποθετείται παράλληλα με την αλληλουχία της πρωτεΐνης. Κάτι τέτοιο δεν είχε παρατηρηθεί στο Zif268, παρότι στις αντίστοιχες θέσεις οι πλευρικές αλυσίδες ήταν στην κατάλληλη απόσταση από τις βάσεις του DNA. Στο μοντέλο του CF2, όμως το ασπαρτικό οξύ 15D από τον δάκτυλο 5 βρίσκεται στην κατάλληλη θέση και έχει την κατάλληλη γεωμετρία για να αλληλεπιδράσει με την αντίστοιχη αδενίνη. Επίσης πειράματα μεταλλαξογένεσης (Πιν. 3) ενισχύουν την υπόθεση ότι το συγκεκριμένο

[8] *Αυτό το πρόγραμμα λειτουργεί ως εξής: Τα δομικά δεδομένα των λυμένων πρωτεϊνών που βρίσκονται στη βάση δεδομένων Brookhaven είναι κωδικοποιημένα υπό μορφή αποστάσεων μεταξύ των ατόμων. Επιλέγοντας λοιπόν μια συγκεκριμένη πρωτεϊνική αλληλουχία το πρόγραμμα βρίσκει ποιά άλλα κομμάτια από λυμένες πρωτεΐνες έχουν παρόμοιες αποστάσεις μεταξύ των ατόμων τους, και παρουσιάζει τα περισσότερο ομόλογα προς αυτή.*

αμινοξύ αλληλεπιδρά με την βάση του DNA. Όπως θα φανεί και από την στατιστική παρακάτω, ο ρόλος αυτής της θέσης είναι μάλλον διπλός, ανάλογα με το τι αμινοξέα υπάρχουν στις άλλες θέσεις της έλικας. Στα πειράματα μεταλλαξογένεσης αυτό που ‘εγινε ήταν το εξής: Για όποιο αμινοξύ στις κρίσιμες θέσεις της έλικας μεταλλάχθηκε (δες Πιν. 3) παράλληλα μετρήθηκε και αν υπήρχε διαφορά στην συγγένεια με την συγκεκριμένη αλληλουχία DNA που δένεται η πρωτεΐνη CF2.

Το μοντέλο μετά από την διεργασία της ελαχιστοποίησης της ενέργειας έδειξε ελάχιστες αλλαγές, όπως ήταν αναμενόμενο. Η γεωμετρία για τα τρία ιόντα του Zn^{+2} ήταν αποδεκτή με τα χαρακτηριστικά της να φαίνονται στον Πιν. 4. Οι αποστάσεις των πλευρικών αλυσίδων του δακτύλου 5 από τις αντίστοιχες βάσεις του DNA, ήταν μεταξύ 2.5 και 3.2 Angstroem. Οι χαρακτηριστικοί υδρογονικοί δεσμοί προς το σκελετό του DNA διατηρήθηκαν κατά αναλογία με το μοντέλο του Zif268. Επίσης το διάγραμμα Ramachadran για την πρωτεΐνη είναι αρκετά καλό και φαίνεται στην Εικ. 13. Εδώ αναφέρουμε ότι οι δίεδρες γωνίες γωνίες των αμινοξέων 141 και 169 και στο μοντέλο του Zif268 βρίσκονται στην ίδια περοχή του διαγράμματος. Μετά από ανάλογη διεργασία, το μοντέλο της πρωτεΐνης με την αντίστοιχη αλληλουχία του πειραματικά λυμένου DNA έδειξε ελάχιστες διαφορές με το παραπάνω μοντέλο του CF2. Συγκεκριμένα η μέση διαφορά (*rms deviation*) μεταξύ των συντεταγμένων όλων των ατόμων τους ήταν 0.155 Angstroem.

Πίνακας 3. Αποτελέσματα πειραμάτων μεταλλαξογένεσης

Θέση	Αμινοξύ	Αποτέλεσμα
Δάκτυλος 4		
13	*Q*	(-)
14	*S*	(+)
15	*N*	(+)
16	*T*	(+)
17	*L*	
18	*K*	(+)
19	*Q*	(+)
Δάκτυλος 5		
13	*V*	(+)
14	*K*	(+)
15	*D*	(+++)
16	*Y*	(+++)
17	*L*	
18	*T*	(+)
19	*K*	(+++)
Δάκτυλος 6		
13	*Q*	(+++)
14	*R*	(+)
15	*S*	(+++)
16	*A-S*	(!)
17	*L*	
18	*T*	(+)
19	*V*	(-)

Συμβολισμοί

(+++) Ελλάτωση στην συγγένεια μεγαλύτερη από 20 φορές
(+) Ελλάτωση στην συγγένεια από 1.5 έως 3 φορές περίπου
(-) Μηδενική ελλάτωση συγγένειας
(!) αύξηση στην συγγένεια περίπου 1.5 φορές

Ας σημειωθεί ότι όλες οι μεταλλάξεις έγιναν προς αλανίνη εκτός από την αλανίνη στην θέση d του δακτύλου 6, που αντικαταστάθηκε από σερίνη.

Πίνακας 4. Γεωμετρία σύνδεσης των ιόντων Zn στα μοντέλα CF2 και Zif268

Μοντέλο	Γωνίες	Αποκλίσεις από το ιδεατό (109.4^{O})
CF2		
δάκτυλος 6		
	102.3	7.1
	132.4	23.0
	98.9	10.5
	135.1	25.7
δάκτυλος 5		
	103.5	5.1
	118.3	8.9
	105.3	4.1
	101.1	8.3
δάκτυλος 4		
	113.4	4.0
	120.2	10.8
	97.8	11.6
	101.1	8.3
Zif268		
δάκτυλος 1		
	108.7	0.7
	111.5	2.1
	116.3	6.9
	103.4	5.0
δάκτυλος 2		
	110.1	0.7
	115.7	6.3
	116.4	7.0
	98.76	10.6
δάκτυλος 3		
	113.4	4.0
	110.8	1.4
	102.7	6.7
	103.4	4.0

Οι παραπάνω γωνίες αναφέρονται σε όλες τις περιπτώσεις με την ακόλουθη σειρά: i) Νε-Zn+2-Sg ii) Sg-Zn+2-Sg iii) Sg- Zn+2-Νε και Ιγ) Νε-Zn+2-Νε.Τα Νε και Sg είναι άτομα αζώτου και θείου από τις δύο ιστιδίνες και κυστεινες αντίστοιχα που δένονται με τον Zn+2. Εδώ αναφέρουμε ότι στον δάκτυλο 6 του CF2 παρατηρούνται λίγο μεγαλύτερες αποκλίσεις αφού έχει γίνει η αλλαγή που προαναφέραμε, σε σχέση με το δομικό πρότυπο του Zif68. Στο πρόγραμμα που χρησιμοποιήσαμε για ελαχιστοποίηση της ενέργειας δεν υπήρχε η δυνατότητα να βελτιωθεί η γεωμετρία σύνδεσης του Zn+2.

III.II ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

III.II.1 Στατιστική των αμινοξέων που αλληλεπιδρούν ειδικά με το DNA

Με το πρόγραμμα που περιγράψαμε στο προηγούμενο κεφάλαιο δημιουργήσαμε μια βάση δεδομένων από 577[9] δακτύλους ψευδαργύρου με την παρακάτω αλληλουχία:

C2X**C**12X**H**3X**H**

Για όλες τις αλληλουχίες υπολογίσαμε τις συχνότητες **R**(r,p), των αμινοξέων που καταλαμβάνουν τις τρείς "κρίσιμες" θέσεις, **13,16,19,** που αλληλεπιδρούν ειδικά με το DNA. Τα ποσοστά φαίνονται στην Εικ. 14. Παρατηρούμε ότι η μεγάλη πλειοψηφία των αμινοξέων, σε αυτές τις θέσεις, είναι υδρόφιλα όπως, η αργινίνη, η λυσίνη, το γλουταμικό οξύ, η γλουταμίνη, η ασπαραγίνη, κ.λ.π. που μπορούν δηλαδή να σχηματίσουν δεσμούς υδρογόνου με τις βάσεις του DNA. Μόνο στην θέση **19** παρατηρείται ένα υδρόφοβο αμινοξύ η βαλίνη, σε ποσοστό 12%. Σ' αυτές τις θέσεις λοιπόν παρατηρούμε κύρια υδρόφιλα αμινοξέα που μπορούν να αλληλεπιδράσουν με τις βάσεις του DNA, όπως έχει βρεθεί και πειραματικά (*Jacobs 1992, Pavletich and Pabo 1991).*

Εξετάζοντας τα ζευγάρια των τριών αυτών θέσεων προσπαθήσαμε να βρούμε ποιά προτιμούν να συνδυάζονται μεταξύ τους. Αυτό επιτεύχθηκε υπολογίζοντας την σχετική συχνότητα (**R.O.**) για ένα συγκεκριμένο ζευγάρι αμινοξέων καθώς και την τυπική τους απόκλιση. Όσο μεγαλύτερο είναι το νούμερο της σχετικής συχνότητας από την μονάδα και όσο μικρότερο είναι το αντίστοιχο της απόκλισης, τόσο για αυτό το ζευγάρι μεγαλώνει η πιθανότητα να μην παρατηρείται κατά τύχη. Για παράδειγμα αν ένα ζευγάρι περιμένουμε να το παρατηρήσουμε κατά τύχη 1/6 φορές και το παρατηρούμε 1/2, με το αντίστοιχο σ να είναι μικρό, τότε λέμε ότι το γεγονός δεν είναι τυχαίο.

[9] *Οι 60 αλληλουχίες του τύπου C2xC12XH4XH μελετήθηκαν ξεχωριστά και τα αποτελέσματα αναφέρονται παρακάτω.*

Προσπαθώντας να αξιολογήσουμε τα δεδομένα διαλέξαμε μόνο τα ζεύγη που ο λόγος της σχετικής τους συχνότηταςπρος την τυπική τους απόκλιση είναι μεγαλύτερος απο 2.5. Αυτό είναι ένα όριο που επιλέξαμε εμπειρικά εξετάζοντας ένα εύρος τιμών, αφού κάτω από αυτό το όριο παρατηρούσαμε ζευγάρια που εμφανιζόταν ελάχιστες φορές. Στην Εικ. 15 φαίνονται τα αντίστοιχα ζευγάρια και οι τιμές τους για τις παραπάνω ποσότητες βρίσκονται στον Πιν. 5

Πίνακας 5
Σχετικές συχνότητες (R.O.) και ο λόγος (R.O./e.s.d's) για τα ζεύγη 13-16, 16-19 και 13-19

Ζευγάρια	**Σχετική συχνότητα (R.O.)**	**(R.O./e.s.d's)**
13-16		
RE	2.03	3.17
QH	1.56	3.18
QN	1.55	4.30
QT	1.45	3.15
QD	1.44	2.82
RH	1.39	3.23
HS	1.39	2.78
RN	1.27	3.34
13-19		
NR	2.83	3.04
RE	2.56	3.03
QL	1.78	2.82
SR	1.78	2.70
HK	1.76	2.93
QT	1.48	3.02
QQ	1.45	3.54
RK	1.43	3.18
RV	1.25	3.12
16-19		
SN	2.90	2.68
EQ	2.82	3.03
EK	2.73	3.29
DK	2.46	2.64
NT	1.96	2.80
SR	1.49	3.46
NE	1.42	3.30
HR	1.35	3.29

Πίνακας 6

Σχετικές συχνότητες (R.O.) και ο λόγος (R.O./e.s.d΄s) για τις τριάδες των θέσεων 13-16-19

Τριάδα 13-16-19	**Σχετική συχνότητα (R.O.)**	**R.O./e.s.d's**
KSN	38.0	2.2
DEQ	24.0	2.2
REQ	9.9	2.2
REK	7.9	2.3
HEK	6.5	2.4
QNT	6.3	2.4
QEK	5.5	2.0
QDK	5.4	1.9
QHQ	4.5	2.3

Από την στατιστική για την τριπλέτα **13-16-19** αναφέρουμε, με το ίδιο σκεπτικό εκείνες με μεγαλύτερη σχετική συχνότητα (**R.O.**) και με την μικρότερη σχετικά τυπική απόκλιση (Πίν. 6). Η σχετική συχνότητα, όπως έχει οριστεί από την (11) εξαρτάται και από τις συχνότητες κάθε αμινοξέος χωριστά. Για παράδειγμα οι τριπλέτες *KSN*, και *DEQ* έχουν σχετικά υψηλές τιμές **R.O**. γιατί τα αντίστοιχα αμινοξέα στις θέσεις **13, 16** και **19** έχουν χαμηλές σχετικές συχνότητες. Ενώ οι υπόλοιπες έχουν αμινοξέα που παρατηρούνται συχνώτερα στις θέσεις **13, 16** και **19** και αυτό ελαττώνει την τιμή της σχετικής τους συχνότητας. Είναι όμως εξίσου σημαντικές αφού ο λόγος τους **R.O./e.s.d΄s** είναι υψηλός. Όπως παρατηρούμε όλες οι τριπλέτες εκτός από μία, την *KSN,* περιλαμβάνουν τουλάχιστον δύο πλευρικές αλυσίδες με παρόμοιο μήκος. Αυτό είναι πολύ σημαντικό, και ενισχύει άμεσα την υπόθεση ότι σχετικά με το μήκος των αμινοξέων οι δάκτυλοι ψευδαργύρου τοποθετούνται ανάλογα προς το DNA. Κάτι τέτοιο παρατηρήσαμε και στο μοντέλο του CF2 που χρειάστηκε σχετική μετατόπιση των δακτύλων 4 και 6, προς το DNA. Επίσης υπάρχουν ορισμένα ζευγάρια και από τους τρείς συνδυασμούς που δεν απαντώνται στις τριπλέτες **13-16-19**. Για παράδειγμα τα ζεύγη *RN, HS, HQ* για τις θέσεις **13-16**, τα *QT, RE, HK*, για τις **13-19** και τα *NE, SR, HR*, για τις **16-19** είναι μερικά από αυτά. Αυτό συμβαίνει ίσως γιατί όταν δύο ζεύγη συνδυάζονται, από μόνα τους, δεν είναι αναγκαίο να συνδυάζονται και μεταξύ τους.

III.II.2 Στατιστική για τις υπόλοιπες θέσεις της έλικας

Η θέση **14** καταλαμβάνεται κυρίως από τρία αμινοξέα, λυσίνη, σερίνη, αργινίνη, και από προλίνη σε ποσοστό 7% (Εικ. 16). Ας σημειωθεί ότι είναι η μόνη θέση, της έλικας που συναντάμε αυτό το αμινοξύ σε τόσο μεγάλο ποσοστό. Στις υπόλοιπες θέσεις τα ποσοστά του κυμαίνονται περίπου σε 1%-2%. Αυτό είναι αναμενόμενο αφού η προλίνη δεν συναντάται συχνά μέσα σε έλικες, γιατί τις τσακίζει. Συνεπώς θα ήταν μάλλον απίθανο να την συναντούσαμε σε υψηλά ποσοστά στις υπόλοιπες θέσεις της έλικας (Richardson & Richardson). Η θέση **14** όμως βρίσκεται μόλις στην αρχή της έλικας σε μια θέση συμβατή για αυτήν, χωρίς να προκαλεί κάποιο ιδιαίτερο τσάκισμα (Richardson & Richardson).

Είναι αξιοσημείωτο ότι όταν υπάρχει προλίνη σ' αυτήν την θέση, στις υπόλοιπες συναντούμε συγκεκριμένα αμινοξέα. Η στατιστική έδειξε σγκεκριμένα ότι στις θέσεις **13**, **16** και **19** προτιμούνται τα R, E, και K αντίστοιχα όταν στη θέση 14 υπάρχει προλίνη (δες Πιν. 7 στο ΠΑΡΑΡΤΗΜΑ). Επίσης ένας δάκτυλος ψευδαργύρου από τον παραπάνω πίνακα αναφέρεται ότι δένεται στην αλληλουχία DNA 5΄-GGGACTTTCC-3΄που έχει βρεθεί σε υποκινητές (*promotors*) από ιούς, όπως ο ιός ανθρώπινης ανοσοεπάρκειας (*HIV-1*), ο *SV40,* και ο *CMV*. Ο λόγος που η προλίνη συνδυάζεται με συγκεκριμένες πλευρικές αλυσίδες στις θέσεις της έλικας, είναι ίσως γιατί δημιουργεί

μια συγκεκριμένη γεωμετρία στην περιοχή.

Η στατιστική για τα ζεύγη της θέσης **14** έδειξε ότι τα προτιμητέα, είναι με αμινοξέα όπως η λυσίνη η αργινίνη και σερίνη. Στο μοντέλο του Zif268 η σερίνη στο δάκτυλο 3 που βρίσκεται σε αυτήν τη θέση σχηματίζει δεσμούς υδρογόνου με το σκελετό της δευτερεύουσας αλυσίδας του DNA, και στους υπόλοιπους δακτύλους ανάλογα αμινοξέα σ' αυτήν την θέση σχηματίζουν δεσμούς με το σκελετό του DNA, έμμεσα μέσω του διαλύτη (*solvent*). Στο μοντέλο του CF2 η λυσίνη του δακτύλου 5 συμπεριφέρεται ανάλογα. Αυτή η θέση λοιπόν είναι σε θέση να αλληλεπιδράσει με τον σκελετό του DNA, είτε άμεσα είτε μέσω του διαλύτη. Όπως φαίνεται και από την Εικ. 16 η λυσίνη και η αργινίνη κυριαρχούν σε αυτήν την θέση που λόγω του θετικού τους φορτίου, είναι χαρακτηριστικά αμινοξέα για τέτοιου είδους αλληλεπιδράσεις (Richardson & Richardson).

Η θέση **15** καταλαμβάνεται σε ποσοστό περίπου 60% από σερίνη (Εικ. 17). Οι διπλέτες με την θέση **13** δείχνουν τα εξής:

Ζεύγος 13-15	**R.O.**	**R.O./e.s.d΄s**
QS	1.1	6.8
RS	1.1	5.2

Εκτός από την ζευγάρι **RD** που δείχνει υψηλή προτίμηση για αυτές τις θέσεις, υπάρχουν και μερικοί άλλοι συνδυασμοί που προτιμούνται, όπως το ζευγάρι **QS** ή το **RS**. Ο λόγος που η τιμή της σχετικής τους συχνότητας είναι μικρός είναι γιατί η σερίνη είναι κατά 60% συντηρημένη, στην θέση **15** και αυτό μειώνει σημαντικά το αντίστοιχο νούμερο, για τους λόγους που έχουμε προαναφέρει. Οι τιμές όμως του λόγου **R.O./e.s.d.'s,** είναι αρκετά ψηλές κάτι που είναι ενδεικτικό. Στο μοντέλο του Zif268 μετρήσαμε την απόσταση μεταξύ του $\mathbf{C_\gamma}$ του ασπαρτικού οξέος (Asp20) και του $\mathbf{N_\alpha}$ της αργινίνης (Arg20) που είναι 3.7 Angstroem. Το οξυγόνο, της πλευρικής αλυσίδας της σερίνης είναι πιο ευέλικτο σ' αυτήν την θέση και μάλλον θα μπορεί να δημιουργήσει δεσμό υδρογόνου με το $\mathbf{N_\alpha}$ της αργινίνης. Γενικά κάτι τέτοιο έχει παρατηρηθεί αρκετές φορές σε λυμένες πρωτεΐνες (Richardson & Richardson). Με ανάλογο τρόπο η σερίνη θα μπορούσε να αλληλεπιδράσει και με την πλευρική αλυσιδα της γλουταμίνης.

Μετά από αυτά συμπεραίνουμε ότι αυτή η θέση πρέπει να παίζει διπλό ρόλο. Να σταθεροποιεί με δεσμούς υδρογόνου μια μεγάλη πλευρική αλυσίδα στην θέση **13**, ή να αλληλεπιδρά ειδικά με βάσεις από την δευτερεύουσα αλυσίδα του DNA. Η πρώτη εκδοχή έχει παρατηρηθεί στο μοντέλο του Zif268, και επίσης υπάρχουν οι ανάλογες ενδείξεις από την στατιστική ανάλυση. Η δεύτερη έχει παρατηρηθεί στο μοντέλο του CF2 στους δακτύλους 5 και 6, και ενισχύεται σημαντικά και από τα πειράματα μεταλλαξογένεσης (Πιν. 3). Επίσης από την στατιστική παρατηρήσαμε κάποιο ποσοστό από δακτύλους που είχε ταυτόχρονα σε δύο από τις τρείς θέσεις **13,16**,**19**, υδρόφοβα αμινοξέα, όπως κυστεΐνη, βαλίνη, αλανίνη, γλυκίνη, λευκίνη κ.λ.π. Τα παραπάνω δεν μπορούν να αλληλεπιδράσουν, λόγω αυτού του γεγονότος με τις βάσεις του DNA. Σ' αυτήν την περίπτωση η θέση **15** καταλαμβάνεται σε ποσοστό 85% από υδρόφιλες πλευρικές αλυσίδες, όπως η σερίνη, η ασπαραγίνη, η τυροσίνη, κ.λ.π. Κάτι ανάλογο συμβαίνει και στον δάκτυλο 6 του CF2 όπου οι θέσεις **16** και **19** καταλαμβάνονται από αλανίνη και βαλίνη αντίστοιχα, ενώ η θέση **15** από σερίνη που βάσει του μοντέλου αλληλεπιδρά ειδικά με το DNA. Κάτι τέτοιο υποδεικνύεται και από τα αποτελέσματα της μεταλλαξογένεσης (Πιν. 3). Όλα αυτά αναδεικνύουν τον σημαντικό ρόλο αυτής της θέσης στη αναγνώριση του DNA στόχου.

Η θέση **17** είναι πολύ καλά συντηρημένη σε ποσοστό 89% από λευκίνη και 5% από φαινυλαλανίνη (Εικ. 18). Αυτό συμβαίνει ίσως γιατί εξυπηρετεί δομικούς σκοπούς δημιουργώντας ένα υδρόφοβο πυρήνα στο εσωτερικό των δακτύλων *(Pavletich and Pabo 1991)*. Ένα ερώτημα που ίσως προκύπτει εδώ είναι, γιατι συναντάμε μόνο την λευκίνη σ' αυτήν την θέση, και όχι και κάποιο άλλο υδρόφοβο αμινοξύ; Η απάντηση είναι γιατί μάλλον η λευκίνη για γεωμετρικούς λόγους (για παράδειγμα λόγοι μεγέθους και στερεοδιάταξης, αφού ο χώρος όπου τοποθετείται η συγκεκριμένη πλευρική αλυσίδα, το

υδρόφοβο εσωτερικό ενός δακτύλου ψευδαργύρου δηλαδή είναι συγκεκριμένος) είναι ιδανική για αυτήν την θέση σε αντίθεση με κάποιο άλλο αμινοξύ, όπως για παράδειγμαη ισολευκίνη που έχει διαφορετική στερεοδιάταξη.

Η θέση **18** καταλαμβάνεται σε μεγαλύτερα ποσοστά από την θρεονίνη, την λυσίνη, την ασπαραγίνη, και από μερικά υδρόφοβα αμινοξέα όπως η ισολευκίνη και η βαλίνη (Εικ. 19). Αυτό επίσης παρατηρείται και στην στατιστική για τα ζευγάρια αυτής της θέσης. Ο προσανατολισμός της θέσης **18** είναι προς το διαλύτη και όχι προς το DNA.

Για αυτήν την οικογένεια δακτύλων ψευδαργύρου οι θέσεις **14** ή **18** μάλλον, δεν είναι δυνατόν να αλληλεπιδράσουν με τις βάσεις του DNA ακόμα και αν καταλαμβάνονται από πολύ μακριές αλυσίδες όπως η λυσίνη. Σ' αυτό το συμπέρασμα καταλήξαμε εξετάζοντας προσεκτικά αυτές τις θέσεις στο μοντέλο του Zif268 και του CF2 επίσης.

Εδώ αξίζει, να σημειωθεί ότι, συγκρίνοντας τους παραπάνω αριθμούς συχνοτήτων (occurences) για τις θέσεις 13-19 της α-έλικας με τα αντίστοιχα που παρατηρούνται στις α-έλικες σε διάφορες πρωτεινες (Richardson & Richardson, 1988) παρατηρούμε τα εξής: για τις θέσεις που έχουν καθορισμένο λειτουργικό ρόλο (π.χ. 13, 16, και, 19) οι συχνότητες των αμινοξέων, στις δύο κατανομές δεν συμφωνούν κάτι που είναι και αναμενόμενο αφού σ_ αυτές πρέπει να υπάρχουν συγκεκριμένα αμινοξέα για γα μπορέσουν γα επιτελέσουν τον ρόλο τους. Ανάλογα συμβαίνει και για τις θέσεις 15 ή 14 που εμπλέκονται, έμεσσα στην διεργασία αναγνώρισης του DNA. Για τις θέσεις 17 ή 18, που ο ρόλος τους είναι, περισσότερο δομικός τα ποσοστά συμφωνούν περισσότεροι επίσης το ότι, συναντούμε προλίνη μόνο στην θέση 14 παρατηρείται γενικά στις α-έλικες σε παρόμοια ποσοστά (*Richardson & Richardson, 1988*).

ΙΙΙ.ΙΙ.3 Στατιστική για την υπόλοιπη αλληλουχία των δακτύλων ψευδαργύρου

Αρχικά μελετήσαμε στατιστικά την αλληλουχία του συνδέσμου μεταξύ των δακτύλων, *GEKP*. Παρατηρήσαμε ότι είναι αρκετά καλά συντηρημένη (Εικ. 20). Πιο συγκεκριμένα η γλυκίνη απαντάται σε ποσοστό 75%, το γλουταμικό οξύ σε 72%, η λυσίνη σε 74%, και η προλίνη σε 80%. Υπάρχουν κάποιοι λόγοι για την συντήρηση της συγκεκριμένης αλληλουχίας. Η γλυκίνη βρίσκεται εκεί εξαιτίας του συγκεκριμένου δομικού προτύπου με το οποίο τερματίζεται η α-έλικα, γνωστή και ως "δομή συνδετήρα" (*c-capping*) (*Pavletich and Pabo, 1991*). Η διαμόρφωση αυτή είναι πολύ χαρακτηριστική και απαιτεί στην θέση αυτή γλυκίνη αφού μπορεί να πάρει θετικές τιμές για την δίεδρη γωνία φ (Richardson & Richardson).

Η προλίνη επίσης είναι πολύ καλά συντηρημένη, και ένας πιθανός λόγος για αυτό είναι γιατί δημιουργεί μια *van der Waals* αλληλεπίδραση με την επόμενη πλευρική αλυσίδα που είναι αρωματικός δακτύλιος σε ποσοστό 88% (τυροσίνη ή φαινυλαλλανίνη). Αυτή η αλληλεπίδραση είναι καθοριστική αφού όπως έχει αναφερθεί και στο μοντέλο του Zif268 περιορίζει τις δυνατές διατάξεις στον χώρο, της αρχής του δακτύλου. Τα άλλα δύο αμινοξέα, το γλουταμικό οξύ και η λυσίνη, δεν αλληλεπιδρούν ούτε με το DNA αλλά ούτε και με την πρωτεΐνη (*Pavletitc and Pabo 1991*). Πιθανόν δεν

παίζουν κάποιο ιδιαίτερο δομικό ή λειτουργικό ρόλο.

Συμπληρωματικά εξετάσαμε και τις δυνατές διαμορφώσεις αυτής της αλληλουχίας σε λυμένες πρωτεΐνες στην βάση δεδομένων *Nrl_3d*. Αρχικά ψάξαμε χρησιμοποιώντας την ρουτίνα *FASTA*, του πακέτου προγραμμάτων *GCG* για την αλληλουχία του συνδέσμου *GEKP*. Δεν βρήκαμε αυτούσια την αλληλουχία και ένας λόγος για αυτό είναι ότι δεν είναι καταχωρημένος κανένας δάκτυλος ψευδαργύρου στην παραπάνω βάση δεδομένων. Βρήκαμε όμως ομόλογες αλληλουχίες, όπου ένα ή το πολύ δύο αμινοξέα μεταβάλλονται και τις συγκρίναμε με το κομμάτι του συνδέσμου *GEKP* μεταξύ του δεύτερου και του τρίτου δακτύλου στο μοντέλο του Zif268. Συγκεκριμένα για κάθε μία μετρήσαμε την μέση διαφορά στις συντεταγμένες των ατόμων τους (*rms deviation*). Οι αλληλουχίες και τα αποτελέσματα[10] φαίνονται στον Πιν. 8.

Πίνακας 8. Αποτελέσματα σύγκρισης της αλληλουχίας του συνδέσμου GEKP με τον αντίστοιχο στο μοντέλο του Zif268

Αλληλουχία	**Πρωτεΐνη**	**Στοιχείο δευτερ. δομής**	**Μ. Διαφορά (Α)**
GEKD	4RUBA	Σύνδεει 2 β-κλώνους	1.12
GEKT	ABP	Σύνδεει β-κλώνο με έλικα	2.47
GEKT	ABP	Συνδέει β-κλώνο με έλικα	2.56
GEKD	FCBA1	Συνδέει β-κλώνο με έλικα	2.11
GEKM	1FNR	Συνδέει β-κλώνο με έλικα	1.21
GEKM	2FNR	Συνδέει β-κλώνο με έλικα	1.23
GDKP	1Cy3	Συνδέει 2 β-κλώνους	1.41
GERP	1phh	Συνδέει 2 β-κλώνους	2.28
GERP	2phh	Συνδέει 2 β-κλώνους	2.28
GAKP	2 fcr	Συνδέει έλικα με β-κλώνο	0.70
GQRP	1tpt	Στροφή πριν από β-κλώνο	1.72
GHRP	4ts1b1	Συνδέει έλικα με β-κλώνο	0.99
HEKP	2sodg	Στροφή μετά από β-κλώνο	1.91
HEKP	2sody	Συνδέει 2 β-κλώνους	1.91
LEKP	3gpdr	Συνδέει β-κλώνο με έλικα	2.10
LERP	9xia	Συνδέει έλικα με β-κλώνο	2.48
LERP	1xis	Συνδέει έλικα με β-κλώνο	2.47
WERP	8apia	Συνδέει 2 β-κλώνους	2.64
LERP	2rspb2	Συνδέει 2 β-κλώνους	1.42
DERP	3aat	Συνδέει έλικα με β-κλώνο	1.12
LERP	2mvpb	Συνδέει 2 β-κλώνους	1.46

[10] *Ας σημειωθεί ότι εδώ αναφέρονται οι αλληλουχίες με τη μεγαλύτερη ομολογία ως προς αυτή του συνδέσμου GEKP.*

Στην πρώτη και στη τρίτη περίπτωση δεν υπάρχει η προλίνη και η γλυκίνη αντίστοιχα, ενώ στην δεύτερη υπάρχουν και τα δύο αυτά αμινοξέα. Παρατηρούμε γενικά μεγάλες διαφορές, εκτός από την αλληλουχίες που συνδέουν έλικα με β-κλώνο και περιέχουν την προλίνη και την γλυκίνη. Κάτι τέτοιο φανερώνει τον σημαντικό ρόλο που παίζουν αυτά τα δύο αμινοξέα, σε σχέση με τα δύο υπόλοιπα. Επίσης παρατηρούμε μικρότερες αποκλίσεις στη περίπτωση που το κομμάτι αυτό συνδέει έλικα και β-κλώνο.

Εκτός από την αλληλουχία του συνδέσμου εξετάσαμε στατιστικά και κάποια άλλα αμινοξέα που φαίνεται να είναι καθοριστικά για την λειτουργία ενός δακτύλου ψευδαργύρου. Η θέση 9 που είναι αργινίνη στο μοντέλο του Zif268 (θέση 14 Εικ. 12) και αλληλεπιδρά με τον σκελετό του DNA, παρατηρήσαμε ότι καταλαμβάνεται από λυσίνη σε ποσοστό 84% (Εικ. 21). Όπως έχει ήδη αναφερθεί αυτά τα δύο αμινοξέα είναι αντιπροσωπευτικά για τέτοιου είδους αλληλεπιδράσεις με νουκλεϊκά οξέα (Richardson & Richardson). Στο μοντέλο του CF2 η λυσίνη (θέση 40 Εικ. 12), που βρίσκεται στον δάκτυλο 5 αλληλεπιδρά με τον σκελετό του DNA με τρόπο ανάλογο όπως η αντίστοιχη αργινίνη του Zif268. Επίσης η θέση 11, που στο μοντέλο του Zif268 είναι φαινυλαλανίνη (θέση 16 Εικ. 12), είναι πολύ καλά συντηρημένη σε ποσοστό 94% (Εικ. 21). Αυτό συμβαίνει γιατί μάλλον παίζει κάποιο δομικό ρόλο συμμετέχοντας στον σχηματισμό ενός υδρόφοβου πυρήνα στο εσωτερικό των δακτύλων.

Συμπερασματικά θα λέγαμε ότι τα αμινοξέα που παίζουν δομικό ρόλο είναι καλά συντηρημένα. Αντίθετα αυτά που αλληλεπιδρούν ειδικά με το DNA, αυτά που βρίσκονται στις στροφές ή αυτά που τοποθετούνται προς το διαλύτη παρουσιάζουν πολυμορφία. Αυτό φαίνεται και στην Εικ. 22 όπου έχει υπλογιστεί η πολυμορφικότητα για κάθε αμινοξύ.

ΙΙΙ.ΙΙ.3 Στατιστική για τις υπόλοιπες οικογένεις των δακτύλων ψευδαργύρου

Κατά την διάρκεια του ψαξίματος της βάσης δεδομένων (*swiss database*) για αλληλουχίες δακτύλων ψευδαργύρου βρήκαμε αποκλίσεις από την αλληλουχία **C**2X**C**12X**H**3X**H**, όπως φαίνεται παρακάτω:

Αριθμός	**Αλληλουχία**
καμμία	**C**2X**C**13X**H**3X**H**
καμμία	**C**2X**C**12X**H**6X**H**
4	**C**2X**C**11X**H**3X**H**
3	**C**2X**C**12X**H**5X**H**
2	**C**2X**C**11X**H**4X**H**
1	**C**4X**C**12X**H**3X**H**
60	**C**2X**C**12X**H**4X**H**[11]

Ολόκληρες οι παραπάνω αλληλουχίες φαίνονται στον Πιν. 9 στο ΠΑΡΑΡΤΗΜΑ. Το γεγονός αυτό, ότι οι εξαιρέσεις από ένα συγκεκριμένο τύπο αλληλουχίας είναι λίγες, υποδηλώνει ότι οι δάκτυλοι ψευδαργύρου δεν παρουσιάζουν πολλές αποκλίσεις μεταξύ τους. Από τα αποτελέσματα της στατιστικής προκύπτει ότι μάλλον δεν διαφέρουν πολύ και στον τρόπο λειτουργίας τους εκτός ίσως ορισμένων επιμέρους αλλαγών, όπως αυτές που παρατηρήθηκαν στο μοντέλο του CF2 σε σχέση με το μοντέλο του Zif268. Η τελευταία οικογένεια, που αναφέρεται παραπάνω είναι η μόνη σημαντική εξαίρεση και για αυτό την μελετήσαμε και στατιστικά. Υπάρχουν μερικές ομοιότητες και διαφορές από την μεγάλη οικογένεια. Για παράδειγμα οι θέσεις **13,16,19** καταλάμβανονται και αυτές κυρίως από υδρόφιλα αμινοξέα αλλά όχι ακριβώς από τα ίδια. Η θέση **15,** για παράδειγμα καταλάμβανεται από σερίνη σε ποσοστό 68%, και η θέση **17** από λευκίνη κατά 71%, αντίστοιχα.

Για τις υπόλοιπες θέσεις υπάρχουν επίσης κάποιες μικρές διαφορές. Η θέση **14** καταλαμβάνεται επίσης από προλίνη σε ποσοστό 23% και παρουσιάζει προτίμηση για συγκεκριμένα αμινοξέα, στις υπόλοιπες θέσεις (δες Πίν. 10 στο ΠΑΡΑΡΤΗΜΑ). Για τις υπόλοιπες συντηρημένες θέσεις απαντώνται και εδώ ίδια αμινοξέα, αλλά σε λίγο διαφορετικά ποσοστά. Τα γενικά χαρακτηριστικά, δηλαδή παραμένουν τα ίδια. Από το μοντέλο του CF2 ο δάκτυλος 6 που ανήκει σ' αυτήν την ομάδα, δείχνει να συμπεριφέρεται με τρόπο ανάλογο με τους υπόλοιπους.

[11] *Οι παραπάνω αλληλουχίες βρίσκονται στον Πίνακα 11 στο ΠΑΡΑΡΤΗΜΑ*

ΚΕΦΑΛΑΙΟ IV

ΣΥΜΠΕΡΑΣΜΑΤΑ

VI. Συζήτηση και μελλοντικά σχέδια

VI.I Συζήτηση

Από το μοντέλο του CF2 και την στατιστική μελέτη ενός μεγάλου αριθμού αλληλουχιών φαίνεται ότι η αλληλεπίδραση, των δακτύλων ψευδαργύρου του τύπου 2**C**2**H**, με το DNA ακολουθεί κάποιους γενικούς κανόνες. Κάθε δάκτυλος αποτελείται από μία αντιπαράλληλη β-πτυχωτή επιφάνεια και μια έλικα, που είναι κυρίως υπεύθυνη για την αλληλεπίδραση με τις βάσεις του DNA. Αυτό το μοτίβο (*motif*) φαίνεται να είναι καλά συντηρημένο με δεδομένη γεωμετρία για το Zn^{+2}, και τα αμινοξέα που παίζουν δομικό ρόλο να συντηρούνται σε μεγάλα ποσοστά. Για παράδειγμα αναφέρουμε τις δύο κυστεΐνες και τις δύο ιστίδινες που δεσμεύουν το ιόν του Zn^{+2}, μερικά άλλα όπως στις θέσεις 9 και 11, κ.λ.π. Μερικές εξαιρέσεις από τον κανόνα, όπως ένα ή δύο αμινοξέα παραπάνω μεταξύ των δύο κυστεϊνών ή ιστιδινών του Zn^{+2}, αντιμετωπίζονται με κάποιες τοπικές δομικές αλλαγές, όπως για παράδειγμα η αλλαγή της διαμόρφωσης ενός βρόγχου (*loop*) κ.λ.π.

Για τις θέσεις **13,14,15,16,17,18,19** θα λέγαμε ότι για συγκεκριμένη γεωμετρία, οι **13, 16,** και **19** είναι σε θέση να αλληλεπιδράσουν ειδικά με το DNA (*Pavletich and Pabo 1991, Jacobs 1992*). Από την στατιστική ανάλυση προκύπτει ότι παρουσιάζουν μεγάλη διαφοροποίηση και καταλαμβάνονται κυρίως από υδρόφιλα αμινοξέα, όπως **Q,R,N,K,H**, που πράγματι μπορούν να σχηματίσουν δεσμούς υδρογόνου με τις βάσεις του DNA. Η θέση **14** φαίνεται ότι αλληλεπιδρά με το σκελετό της δευτερεύουσας αλυσίδας (*secondery strand*) και χρειάζεται ανάλογα με την σχετική θέση προς το DNA, μια μακριά ή μια κοντή πλευρική αλυσίδα. Η θέση **15** μπορεί να σταθεροποιήσει ένα μεγάλο αμινοξύ στην θέση **13** (Pavletich and Pabo 1991), ή να αλληλεπιδράσει ειδικά με βάσεις από την δευτερεύουσα αλυσίδα του DNA. Αυτές οι δύο υποθέσεις ενισχύονται από την στατιστική, από το μοντέλο του CF2 για τους δακτύλους 5 και 6, και επίσης από τα πειράματα μεταλλαξογένεσης. Η θέση **17** είναι συντηρημένη σε ποσοστό 89% αφού μάλλον παίζει δομικό ρόλο στον σχηματισμό ενός υδρόφοβου εσωτερικού για τον δάκτυλο (*Pavletich & Pabo 1991*). Η θέση **18** τοποθετείται προς το εξωτερικό και σχηματίζει δεσμούς υδρογόνου με τον διαλύτη.

Μια ερώτηση που ίσως γεννάται εδώ είναι: αν θα ήταν δυνατόν να στραφεί ένας δάκτυλος κατά μια γωνία, ως προς τον άξονα της έλικας ώστε να μπορούν να αλληλεπιδράσουν διαφορετικές θέσεις με τις βάσεις του DNA. Κάτι τέτοιο δεν φαίνεται πολύ πιθανό για πολλούς λόγους, όπως αυτοί που αναφέρουμε παρακάτω:

i) Οι θέσεις που παίζουν καθοριστικό ρόλο στην αλληλεπίδραση φαίνεται ότι

είναι καθορισμένες (π.χ. 13.16.19 ή και 15).

ii) Το ιόν του Zn^{+2}, που δεσμεύεται από δύο αμινοξέα της έλικας και δύο από την β-πτυχωτή επιφάνεια, δεν επιτρέπει μια στροφή μόνο της έλικας.

iii) Υπάρχουν χαρακτηριστικοί δεσμοί υδρογόνου από συντηρημένα αμινοξέα προς το σκελετό του DNA, όπως για παράδειγμα μια ιστιδίνη του Zn^{+2}, που είναι δεσμευτικοί για τέτοιου είδους μετακινήσεις.

iv) Οι θέσεις της έλικας που θα ήταν σε θέση να αλληλεπιδράσουν ειδικά με το DNA, θα άλλαζαν στις **15**,**17** ή **18** για παράδειγμα. Κάτι τέτοιο όμως δεν θα είχε νόημα αφού αυτές οι θέσεις δεν παρουσιάζουν πολυμορφία και επίσης δεν καταλαμβάνονται μόνο από υδρόφιλα αμινοξέα. Φαίνεται λοιπόν ότι υπάρχουν αρκετοί περιορισμοί για μια τέτοια περιστροφή που θα άλλαζε τα γεωμετρικά δεδομένα της αλληλεπίδρασης.

VI.II Μελλοντικά σχέδια

Πέρα από την παρούσα εργασία θα μπορούσαν να γίνουν περαιτέρω πειράματα πάνω σ' αυτό το αντίκειμενο.

Για παράδειγμα, θα μπορούσαν να γίνουν επιπλέον πειράματα μεταλλαξογένεσης για την πλειρέστερη κατανόηση του ρόλου ορισμένων αμινοξέων στην διεργασία της αναγνώρισης του DNA στόχου. Επίσης θα προτείναμε και κάποιες άλλες κρυσταλλογραφικές μελέτες ανάλογες με αυτή του Zif268, ή παρόμοιες μελέτες με N.M.R και σε κάποια άλλα συμπλοκα πρωτεϊνών με DNA.

Κάτι που δεν είναι απόλυτα ξεκάθαρο είναι η λειτουργική σημασία μερικών δακτύλων ψευδαργύρου, σε μια αλληλουχία τους με μεγάλο αριθμό, π.χ. 9. Μέχρι τώρα πιστεύεται ότι αφού κάποια από αυτά δεν μπορούν να τοποθετηθούν στην κύρια αύλακα του DNA δεν θα παίζουν απαραίτητα και κάποιο λειτουργικό ρόλο. Οι ενδείξεις για κάτι τέτοιο δεν είναι πολλές και μια κρυσταλλογραφική μελέτη σε τέτοιες αλληλουχίες δακτύλων ψευδαργύρου μαζί με το DNA θα ήταν πολύ διαφωτιστική.

Κάτι που είναι εξίσου σημαντικό είναι να γίνουν περαιτέρω πειράματα για την αποκρυπτογράφηση του κώδικα αναγνώρησης, πρωτεΐνης και DNA. Αυτό μπορεί να γίνει με τις υπάρχουσες τεχνικές της μοριακής βιολογίας και της βιοχημείας. Αν επιτευχθεί κάτι τέτοιο θα επιφέρει επαναστατικές αλλαγές στους τομείς πολλών επιστημών, όπως η μοριακή βιολογία, η βιοτεχνολογία, η ιατρική κ.λ.π., και στις εφαρμογές τους σε τομείς της σύγχρονης ζωής.

Με γνωστά τα στατιστικά δεδομένα για την οικογένεια δακτύλων ψευδαργύρου του τύπου **C C/H H** είναι δυνατόν πλέον, να αναγνωρίσει κανείς με μεγάλη βεβαιότητα αν μια αλληλουχία ανήκει σ' αυτήν ή όχι. Μέχρι τώρα δεν ήταν γνωστά πολλά

δεδομένα για την αλληλουχία αυτών των περιοχών. Μελλοντικά σε συνδυασμό με την ύπαρξη του κώδικα αναγνώρησης του DNA θα είναι δυνατόν να ελέγχουμε την γονιδιακή ρύθμιση σε κάποιο βεθμό. Κάτι τέτοιο θα δώσει απαντήσεις σε πολλά βιολογικά προβλήματα, και ταυτόχρονα θα εφαρμοστεί σε πολλούς τομείς της σύγχρονης τεχνολογίας. Τα παραδείγματα που θα μπορούσαμε να αναφέρουμε είναι ακόμα περισσότερα αυτά που δεν μπορούμε να διανοηθούμε προς το παρόν. Αναφέρουμε χαρακτηριστικά την θεραπεία πολλών γεννετικών ασθενειών που οφείλονται στην κακή ή ελαττωματική λειτουργία τέτοιων πρωτεϊνών. Επίσης στον τομέα της βιοτεχνολογίας θα είναι δυνατόν να παρασκευασθούν ευκολώτερα πολλές πρωτεΐνες, σε μεγάλες ποσότητες δύνοντας λύση σε πολλά προβλήματα της τεχνολογίας φαρμάκων.

Κλείνοντας θα λέγαμε ότι τα μονοπάτια που οδηγούν στην πλήρη κατανόηση πολύπλοκων βιολογικών φαινομένων ανοίγονται μέρα με την ημέρα. Απαιτούνται όμως πολλές προσπάθειες ακόμα, για την επίτευξη του τελικού στόχου.

ΠΑΡΑΡΤΗΜΑ

Πινακας 2. Πρωτεΐνες που περιέχουν περιοχές δακτύλων ψευδαργύρου

Egr1_Human P18146 homo sapiens (human). early growth response protein 1 (egr-1) (krox24) (transcription factor)
Mgf1_Mouse P16372 mus musculus (mouse). zinc finger protein mfg1 (fragment). 5/92 169bp
Mgf2_Mouse P16373 mus musculus (mouse). zinc finger protein mfg2 (fragment). 5/92 406bp
Mgf3_Mouse P16374 mus musculus (mouse). zinc finger protein mfg3 (fragment). 5/92 305bp
Mok2_Mouse P24399 mus musculus (mouse). zinc finger protein mok2. 3/92 201bp
Sdc1_Caeel P24349 caenorhabditis elegans. zinc-finger protein sdc-1. 8/92 1,203bp
Vcg3_Npvac P16091 autographa californica nuclear polyhedrosis virus (acmnpv). dna-binding protein (zinc finger
Zf26_Mouse P10076 mus musculus (mouse). zinc finger protein zfp-26 (mkr3 protein) (fragment). 5/92 428bp
Zf27_Mouse P10077 mus musculus (mouse). zinc finger protein zfp-27 (mkr4 protein) (fragment). 5/92 419bp
Zf28_Mouse P10078 mus musculus (mouse). zinc finger protein zfp-28 (mkr5 protein) (fragment). 5/92 384bp
Zf35_Mouse P15620 mus musculus (mouse). zinc finger protein zfp-35. 8/91 580bp
Zf36_Human P16415 homo sapiens (human). zinc finger protein zfp-36 (fragment). 2/91 582bp
Zf37_Mouse P17141 mus musculus (mouse). zinc finger protein zfp-37 (male germ cell specific zinc finger protein
Zf64_Human P15622 homo sapiens (human). zinc finger protein clone 647 (fragment). 11/90 216bp
Zfa_Mouse P23607 mus musculus (mouse). zinc finger autosomal protein. 11/91 742bp
Zfp1_Mouse P08042 mus musculus (mouse). zinc finger protein zfp-1 (mkr1 protein). 5/92 424bp
Zfx1_Mouse P17011 mus musculus (mouse). zinc finger x-chromosomal protein (clone pdp1115). 2/91 799bp
Zfx2_Mouse P17012 mus musculus (mouse). zinc finger x-chromosomal protein (clone pdp1119). 2/91 839bp
Zfx_Human P17010 homo sapiens (human). zinc finger x-chromosomal protein. 2/91 805bp
Zfy1_Mouse P10925 mus musculus (mouse). zinc finger y-chromosomal protein 1. 2/91 782bp
Zfy2_Mouse P20662 mus musculus (mouse). zinc finger y-chromosomal protein 2. 2/91 783bp
Zfy_Human P08048 homo sapiens (human). zinc finger y-chromosomal protein (possible testis determining factor).
Zg16_Xenla P18712 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf16.1 (fragment). 11/9
Zg17_Xenla P18713 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf17.1 (fragment). 11/9
Zg20_Xenla P18714 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf20.1 (fragment). 11/9
Zg26_Xenla P18715 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf26.1 (fragment). 11/9
Zg28_Xenla P18716 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf28.1 (fragment). 11/9
Zg29_Xenla P18717 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf29.1 (fragment). 11/9
Zg32_Xenla P18719 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf32.1 (fragment). 11/9
Zg3_Xenla P18718 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf3.1 (fragment). 11/90
Zg42_Xenla P18720 xenopus laevis (african clawed frog). gastrula zinc finger

protein xlcgf42.1 (fragment). 11/9
Zg44_Xenla P18721 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf44.2 (fragment). 11/9
Zg46_Xenla P18722 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf46.1 (fragment). 11/9
Zg48_Xenla P18723 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf48.2 (fragment). 11/9
Zg49_Xenla P18724 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf49.1 (fragment). 11/9
Zg52_Xenla P18727 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf52.1 (fragment). 11/9
Zg53_Xenla P18728 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf53.1 (fragment). 11/9
Zg57_Xenla P18729 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf57.1 (fragment). 11/9
Zg58_Xenla P18730 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf58.1 (fragment). 11/9
Zg5_Xenla P18725 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf5.1 (fragment). 11/90
Zg5a_Xenla P18726 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf51.1a (fragment). 11/
Zg62_Xenla P18731 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf62.1 (fragment). 11/9
Zg64_Xenla P18732 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf64.1 (fragment). 11/9
Zg66_Xenla P18733 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf66.1 (fragment). 11/9
Zg67_Xenla P18734 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf67.1 (fragment). 11/9
Zg71_Xenla P18736 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf71.1 (fragment). 11/9
Zg7_Xenla P18735 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf7.1 (fragment). 11/90
Zg8_Xenla P18737 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf8.2db (fragment). 11/
Zg9_Xenla P18738 xenopus laevis (african clawed frog). gastrula zinc finger protein xlcgf9.1 (fragment). 11/90
Zk05_Human P17016 homo sapiens (human). zinc finger protein kox5 (fragment). 5/91 56bp
Zk06_Human P17017 homo sapiens (human). zinc finger protein kox6 (fragment). 5/91 56bp
Zk08_Human P17019 homo sapiens (human). zinc finger protein kox8 (fragment). 5/91 56bp
Zk09_Human P17020 homo sapiens (human). zinc finger protein kox9 (fragment). 5/91 56bp
Zk10_Human P17021 homo sapiens (human). zinc finger protein kox10 (fragment). 5/91 56bp
Zk11_Human P17022 homo sapiens (human). zinc finger protein kox11 (fragment). 5/91 56bp
Zk13_Human P17024 homo sapiens (human). zinc finger protein kox13 (fragment). 5/91 56bp
Zk14_Human P17025 homo sapiens (human). zinc finger protein kox14 (fragment). 5/91 56bp
Zk16_Human P17027 homo sapiens (human). zinc finger protein kox16 (fragment). 5/91 56bp
Zk18_Human P17029 homo sapiens (human). zinc finger protein kox18 (fragment). 5/91 56bp
Zk21_Human P17032 homo sapiens (human). zinc finger protein kox21 (fragment). 5/91 56bp
Zk22_Human P17033 homo sapiens (human). zinc finger protein kox22 (fragment). 5/91 56bp

Zk23_Human P17034 homo sapiens (human). zinc finger protein kox23 (fragment). 5/91 56bp
Zk24_Human P17035 homo sapiens (human). zinc finger protein kox24 (fragment). 5/91 56bp
Zk25_Human P17036 homo sapiens (human). zinc finger protein kox25 (fragment). 5/91 56bp
Zk27_Human P17038 homo sapiens (human). zinc finger protein kox27 (fragment). 5/91 56bp
Zk28_Human P17039 homo sapiens (human). zinc finger protein kox28 (fragment). 5/91 50bp
Zk29_Human P17040 homo sapiens (human). zinc finger protein kox29 (fragment). 5/91 56bp
Zk30_Human P17041 homo sapiens (human). zinc finger protein kox30 (fragment). 5/91 52bp
Zkr1_Chick P30373 gallus gallus (chicken). zinc finger protein ckr1. 4/93 509bp
Zn07_Human P17097 homo sapiens (human). zinc finger protein 7 (zinc finger protein kox4). 12/92 686bp
Zn08_Human P17098 homo sapiens (human). zinc finger protein 8 (fragment). 12/92 543bp
Zn10_Human P21506 homo sapiens (human). zinc finger protein 10 (zinc finger protein kox1) (fragment). 12/92 462
Zn11_Human P17013 homo sapiens (human). zinc finger protein 11a (zinc finger protein kox2) (fragment). 12/92 56
Zn12_Human P17014 homo sapiens (human). zinc finger protein 12 (zinc finger protein kox3) (fragment). 12/92 56b
Zn19_Human P17023 homo sapiens (human). zinc finger protein 19 (zinc finger protein kox12) (fragment). 12/92 56
Zn22_Human P17026 homo sapiens (human). zinc finger protein 22 (zinc finger protein kox15) (fragment). 12/92 56
Zn24_Human P17028 homo sapiens (human). zinc finger protein 24 (zinc finger protein kox17) (fragment). 12/92 44
Zn25_Human P17030 homo sapiens (human). zinc finger protein 25 (zinc finger protein kox19) (fragment). 12/92 56
Zn26_Human P17031 homo sapiens (human). zinc finger protein 26 (zinc finger protein kox20) (fragment). 12/92 56
Zn29_Human P17037 homo sapiens (human). zinc finger protein 29 (zinc finger protein kox26) (fragment). 8/92 56b
Zn35_Human P13682 homo sapiens (human). zinc finger protein 35 (zinc finger protein hf.10). 12/92 491bp
Zn40_Human P15822 homo sapiens zinc finger protein 40 (human immunodeficiency virus type i enhancer- binding pr
Zn42_Human P28698 homo sapiens (human). zinc finger protein 42 (myeloid zinc finger 1) (mzf-1). 12/92 485bp
Zn43_Human P28160 homo sapiens (human). zinc finger protein 43. 12/92 803bp
Zn44_Human P15621 homo sapiens (human). zinc finger protein 44 (zinc finger protein kox7) (clone 431) (fragment
Zn46_Human P24278 homo sapiens (human). zinc finger protein 46 (zinc finger protein kup). 12/92 433bp
Znf1_Canal P28875 candida albicans (yeast). zinc finger protein 1. 12/92 388bp
Znfp_Lycva P18541 lymphocytic choriomeningitis virus (strain armstrong). zinc finger protein. 2/91 90bp
Znfp_Lycvp P19326 lymphocytic choriomeningitis virus (strain pasteur). zinc finger protein (fragment). 2/91 51b
Znfp_Lycvt P19325 lymphocytic choriomeningitis virus (strain traub). zinc finger protein (fragment). 2/91 72bp
Zo10_Xenla P18739 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof10 (fragment). 11/90 21
Zo14_Xenla P18740 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof14 (fragment). 11/90 13
Zo15_Xenla P18741 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof15 (fragment). 11/90 14

Zo19_Xenla P18742 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof19 (fragment). 11/90 16
Zo20_Xenla P18744 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof20 (fragment). 11/90 24
Zo22_Xenla P18745 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof22. 11/90 435bp
Zo26_Xenla P18746 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof26 (fragment). 11/90 19
Zo28_Xenla P18747 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof28 (fragment). 11/90 43
Zo29_Xenla P18748 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof29. 11/91 537bp
Zo2_Xenla P18743 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof2 (fragment). 11/90 157
Zo61_Xenla P18750 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof6.1 (fragment). 11/90 2
Zo6_Xenla P18749 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof6 (fragment). 11/90 453
Zo71_Xenla P18751 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof7.1 (fragment). 11/90 8
Zo72_Xenla P18752 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof7.2 (fragment). 11/90 3
Zo84_Xenla P18753 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof8.4 (fragment). 11/90 7
Zo8i_Xenla P18853 xenopus laevis (african clawed frog). oocyte zinc finger protein xlcof8.4i (fragment). 11/90

Πινακας 7. Αλληλουχίες της μεγάλης οικογένειας που περιέχουν προλίνη στη θέση b

ZFA_MOUSE PHKCDMCDKGFIIKPSELKKIIV AAIIKGKK P23607 mus musculus (mouse). zinc finger autosomal protein. 598
ZFA_MOUSE PHICVECGKGFCIIPSELKKHMKIHTGEK P23607 mus musculus (mouse). zinc finger autosomal protein. 484
ZFA_MOUSE PHCCEIICKKGFKRPSEKNQH IMRHIIKEV P23607 mus musculus (mouse). zinc finger autosomal protein. 712
ZFXl_MOUSE PHKCDMCDKGFHRPSELKKHVAAHKGKK P17011 mus musculus (mouse). zinc finger x-chromosomal prote 655
ZFXl_MOUSE PHRCEYCKKGFRKPSEKNQHIMRHHKEV P17011 mus musculus (mouse). zinc finger x-chromosomal prote 769
ZFX1 MOUSE PHICVECGKGFRIIPSELKKHMKIHTGEK P17011 mus musculus (mouse). zinc finger x-chromosomal prote 541
ZFX2 MOUSE PHKCDMCDKGFHKPSELKKHV AAIIKGKK P 17012 mus musculus (mouse) - zinc finger x-chromosomal prote 695
ZFX2_MOUSE PHICVECGKGFRHPSELKKIIMRIHTGEK P17012 mus musculus (mouse). zinc finger x-chromosomal prote 581
ZFX2_MOUSE PHRCEYCKKGFRRPSEKNQHIMRHHKEV P17012 mus musculus (mouse). zinc finger x-chromosomal prote 809
ZFX_HUMAN PHKCDMCDKGFHRPSELKKIHVAAHKGKK P17010 homo sapiens (fan)- zinc finger x-chromosomal prote 661
ZFX HUMAN PHICVECGKGFRHPSELKKHMKIHTGEK P17010 homo sapiens (human)- zinc finger x-chromosomal prote 547
ZFX HUMAN PHRCEYCKKGFRRPSEKNQHIMRHHKEV P17010 homo sapiens (human). zinc finger x-chromosomal prote 775
ZFYl_MOUSE PHSCDFCKKGFRRPSEKNQHIMRHHKVG Pl0925 mus musculus (mouse). zinc finger y-chromosomal prote 753
ZFY l_MOUSE PHKCDMCSKGFHRPSELKKHVATHKSKK P 10925 mus musculus (mouse). zinc finger y-chromosomal prote 639
ZFY l_MOUSE PHICGECGKGFRHPSALKKHIRVHTGEK P 10925 mus musculus (mouse). zinc finger y-chromos omal protein
525 ZFYZ_MOUSE PHICGECGKGFRHPSALKKHIRVHTGEK P20062 mus musculus (mouse). zinc finger y-chromosomal prote 525
ZFY2_MOUSE PHKCDMCSKGFHRPSELKKHV ATHKSKK P20062 mus musculus (mouse). zinc finger y-chromosomal prote 639
ZFY2_MOUSE PHRCDFCKKGFRRPSEKNQHIMRHHKEV P20062 mus musculus (mouse). zinc finger y-chromosomal prote 753
ZFY_HUMAN PHKCEMCEKGFHRPSELKKHV AVHKGKK P08008 homo sapiens (human). zinc finger y-chromosomal prote 657
ZFY_HUMAN PHRCEY CKKGFRRPSEKNQHIMRHHKEV P08008 homo sapiens (human). zinc finger y-chromosomal prote 771
ZFY_HUMAN PHICVECGKGFRYPSELRKHMRIHTGEK P08008 homo sapiens (human). zinc finger y-chromosomal prote 543
Z028 XENLA PHKCNLCDKTFHYPSNLVEHQRTHTGDR P18747 xenopus laevis (african clawed frog). oocyte zinc fin 117
Z071 XENLA PFSCSECGKGFTRPNALIIHHRTHTGEK P18751 xenopus laevis (african clawed frog). oocyte zinc fin 369
Z071 XENLA PFSCFECRKCFSNPSNLARHQMTHTGEK P18751 xenopus laevis (african clawed frog). oocyte zinc fin 594
Z084_XENLA PFSCSECGKCFSTPHVRARHQKTHTGEK P18753 xenopus laevis (african clawed frog). oocyte zinc fin 305
MGFZ_MOUSE PFECKECGKAFLLPSQLNSHKIVHTSKR P 16373 mus musculus (mouse). zinc finger protein mfg2 (fragm 307
ZF36_HUMAN PYKCKLCGKAFVWPSLFHLHERTHTGEK Pl6415 homo sapiens (human). zinc finger protein zfp-36 (fra 225
ZF36_HUMAN PHKCKICGKGFDCPSSVRNIIETTHTGEK Pl6451homo sapiens (human). zinc finger protein zfp-36 (fra 365
ZF36_HUMAN PQKCKICGKAFGCPSLFQRHERTHTGEK homo sapiens (human). zinc finger protein zfp-36 (fra 421

ZK 13_HUMAN ADECKECGNAFSFPSEIRRHKRSHTGEK P 17024 homo sapiens (human). zinc finger protein kox 13 (fra
ZN40_HUMAN KY ICEYCNRACAKPSVLLKHIRSHTGER homo sapiens zinc finger protein 40 (human immunod(:fi
ZN40_HUMAN KYICEECGIRCKKPSMLKKHIRTHTDVR homo sapiens zinc finger protein 40 (human immunodt fi
ZN43 HUMAN PYKCEECGKAFNWPSTLTKHKRIHTGEK homo sapiens (human). zinc finger protein 43. 12/92
ZN43_HUMAN PYKCEECGKAFNWPSTLTKHNRIHTGEK homo sapiens (human). zinc finger protein 43. 12/92
ZN43 HUMAN PYKCEKCGKAFNRPSNLIEHKKIHTGEQ homo sapiens (human). zinc finger protein 43. 12/92
ZN44 HUMAN PHKCTVCGKAFDSPSVFQRHERTHTGEK homo sapiens (human). zinc finger protein 44 (zinc fi

Πινακας 9. Αλληλουχίες με διαφορετική αλληλουχία από την πλειοψηφία των δακτύλων ψευδαργύρου

Αλληλουχίες της μορφής C2XC11XH3H

MGF2_MOUSE PFQCEACGKSLANTLLIHHQKSHSGERPF P16373 mus musculus (mouse). zinc finger protein mfg2 (fragm 280
SDCl_CAEEL MSSCHLCHLPVPNKFLEAHGNVHRGRFRI P24349 caenorhabditis elegans. zinc-finger protein sdc- 1. 81 234
ZG20_XENLA AFSCNLCDKLSIISKLRLHYRVHSGEKPY P18714 xenopus laevis (african clawed frog). gastrula zinc f 395
ZG52_XENLA PFTCPECGKRFSQKSNCWHTEDHTGEKPF P18727 xenopus laevis (african clawed frog). gastrula zinc f 5

Αλληλουχίες της μορφής C2XC11XH4XH

SDCl_CAEEL VVVCFIICGTRCHYTLLHDHLDYCHYWPRN P24349 caenorhabditis elegans. zinc-finger protein sdc-l. 8/ 653
ZN07_HUMAN ISRCQECQKKLSDCLQGKHTNNCHGEKPY P17097 homo sapiens (human). zinc finger protein 7 (zinc fin 222

Αλληλουχίες της μορφής C2XC11XH5XH

ZKRl_CHICK PQRCAECGKAFRAAPPLRRHRRERSHRCGDCGKGF P30373 gallus gallus (chicken). zinc finger protein ckrl. 41 278
ZN40_HUMAN PYPCVTCGFSFKTKSNLYKHKKSHAHTIKLGLVLQ P15822 homo sapiens zinc finger protein 40 (human immunodefi 433
ZN40 HUMAN PYHCTY CNFSFKTKGNLTKHMKSKAHSKKCVDLGI P15822 homo sapiens zinc finger protein 40 (human immunodefi 2114

Αλληλουχίες της μορφής C4XC12XH3XH

EGRl_HUMAN PYACPVESCDRRFSRSDELTRHIRIHTGQ P18146 homo sapiens (human) early growth response protein 1 337

Πινακας 10. Αλληλουχίες της μικρής οικογένειας με προλίνες στη θέση b

ZFA MOUSE THQCRIICDFKIADPFVLSRH lLSVHTKDL P23607 mus musculus (mouse). zinc finger autosomal protein. 626
ZFA MOUSE PHCCEHCKKGFRRPSEKNQHIMRHHKEVG P23607 mus musculus (mouse). zinc finger autosomal protein. 712
ZFXl MOUSE PHRCEYCKKGFRRPSEKNQHIMRHI~KEVG P17011 mus musculus (mouse). zinc finger x-chromosomal prote 769
ZFX l_MOUSE MHQCRHCDFKIADPFVLSRHILSVHTKDL P17011 mus musculus (mouse). zinc finger x-chromosomal prote 683
ZFX2 MOUSE PIIRCEY CKKGFRRPSEKNQHIMRHHKEVG P17012 mus musculus (mouse). zinc finger x-chromosomal prote 809
ZFX2 MOUSE MHQCRHCDFKIADPFVLSRHILSVHTKDL P17012 mus musculus (mouse). zinc finger x-chromosomal prote 723
ZFX HUMAN PHRCEYCKKGFRRPSEKNQHIMRHHKEVG P17010 homo sapiens (human). zinc finger x-chromosomal prote 775
ZFX_HUMAN MHQCRHCDFKIADPFVLSRHILSVHTKDL P 17010 homo sapiens (human). zinc finger x-chromosomal prote 689
ZFYl MOUSE PHSCDFCKKGFRRPSEKNQHIMRHHKVGL Pl0925 mus musculus (mouse). zinc finger y-chromosomal prote 753
ZFY l_MOUSE MHQCRHCDFNSPDPFLLSHHILSAHTKNV P 10925 mus musculus (mouse). zinc finger y-chromosomal prote 667
ZFY2_MOUSE PHRCDFCKKGFRRPSEKNQHIMRHHKEVG P20662 mus musculus (mouse). zinc finger y-chromosomal prote 753
ZFY2_MOUSE MHQCRHCDFNSPDPFLLSHHILSAHTKNV P20062 mus musculus (mouse). zinc finger y-chromosomal prote 667
ZFY_HUMAN MHQCRHCDFKIADPFVLSRHILSVHTKDL P08008 homo sapiens (human). zinc finger y-chromosomal prote 685
ZFY_HUMAN PHRCEYCKKGFRRPSEKNQHIMRHHKEVG P08008 homo sapiens (human). zinc finger y-chromosomal prote 771

Πινακας 11. Αλληλουχίες της μορφής C2XC12XH4XH

ZF28 MOUSE PYECPECGKAFIQNTSLVRHWRYYHTGEK PlO078 mus musculus (mouse). zinc finger protein zfp-28 (mkr 165
ZF36 HUMAN PFDCKECAKTFSSLGNLRRHMAAGDGP P16415 homo sapiens (human). zinc finger protein zfp-36 (fra 197
ZFA_MOUSE MHKCKFCEYETAEQGLLNRIILLAVIISKNF P23607 mus musculus (mouse). zinc finger autosomal protein. 455
ZFA_MOUSE PYECQYCEYRSTDSSNLKTHVKTKlISKEM P23607 mus musculus (mouse). zinc finger autosomal protein. 512
ZF A MOUSE THQCLIICDHKSSNSSDLKRHllSVllTKDY P23607 mus musculus (mouse). zinc finger autosomal protein. 569
ZFA_MOUSE THQCRHCDFKIADPFVLSRHILSVHTKDL P23607 mus musculus (mouse). zinc finger autosomal protein. 626
ZFA MOUSE VYQCEYCDYSTTDASGFKRHVISHITKDY P23607 mus musculus (mouse). zinc finger autosomal protein. 683
ZFA MOUSE PHCCEHCKKGFRRPSEKNQHIMRHHKEVG P23607 mus musculus (mouse). zinc finger autosomal protein. 712
ZFXl_MOUSE MIIKCKFCEYETAEQGLLNRHLLAVHSKNF P17011 mus musculus (mouse). zinc finger x-chromosomal prote 512
ZFXl_MOUSE PYECQYCEYRSADSSNLKTHVKTKHSKEM P17011 mus musculus (mouse). zinc finger x-chromosomal prote 569
ZFXl_MOUSE THQCLHCDHKSSNSSDLKRHIISVHTKDY P17011 mus musculus (mouse). zinc finger x-chromosomal prote 626
ZFX l_MOUSE MHQCRHCDFKIADPFVLSRHlLSVHTKDL P17011 mus musculus (mouse). zinc finger x-chromosomal prote 683
ZFXl_MOUSE VYQCEYCEYSTTDASGFKRHVISIHTKDY P17011 mus musculus (mouse). zinc finger x-chromosomal prote 740
ZFXl_MOUSE PHRCEYCKKGFRRPSEKNQHIMRHHKEVG P17011 mus musculus (mouse). zinc finger x-chromosomal prote 769
ZFX2_MOUSE MHKCKFCEYET AEQGLLNRHLLAVHSKF P17012 mus musculus (mouse). zinc finger x-chromosomal prote 552
ZFX2_MOUSE PYECQYCEYRSADSSNLKTHVKTKHSKEM P17012 mus musculus (mouse). zinc finger x-chromosomal prote 609
ZFX2_MOUSE THQCLHCDHKSSNSSDLKRHIISVHTKY P17012 mus musculus (mouse). zinc finger x-chromosomal prote 666
ZFX2_MOUSE MHQCRHCDFKIADPFVLSRHILSVHTKDL P17012 mus musculus (mouse). zinc finger x-chromosomal prote 723
ZFX2_MOUSE VYQCEYCEYSTTDASGFKRHVISIHTKDY P17012 mus musculus (mouse). zinc finger x-chromosomal prote 780
ZFX2_MOUSE PHRCEYCKKGFRRPSEKNQHIMRHHKEVG P17012 mus musculus (mouse). zinc finger x-chromosomal prote 809
ZFX_HUMAN MHKCKFCEYETAEQGLLNRHLLAVHSKNF P17010 homo sapiens (human). zinc finger x-chromosomal prote 518
ZFX_HUMAN PYQCQYCEYRSADSSNLKTHVKTKHSKEM P17010 homo sapiens (human). zinc finger x-chromosomal prote 575
ZFX_HUMAN THQCLHCDHKSSNSSDLKRHIISVHTKDY P17010 homo sapiens (human). zinc finger x-chromosomal prote 632
ZFX_HUMAN MHQCRHCDFKIADPFVLSRHILSVHTKDL P17010 homo sapiens (human). zinc finger x-chromosomal prote 689
ZFX_HUMAN VYQCEYCEYSITDASGFKRHVISIHTKDY P17010 homo sapiens (human). zinc finger x-chromosomal prote 746
ZFX_HUMAN PHRCEYCKKGFRRPSEKNQHIMRHHKEVG P17010 homo sapiens (human). zinc finger x-chromosomal prote 775
ZFY l_MOUSE TCKCKFCDYETAEQTLLNHHLLVVHRKKF P10925 mus musculus (mouse). zinc finger y-chromosomal prote 496

ZFY1 MOUSE PYFECQYCEYKSADSSNLKTHIKSK I ISKEI P10925 mus musculus (mouse). zinc finger, y-chromosomal prote 553
ZFY1 MOUSE THQCSHCNHKSSNSSDLKRHIISVHTKAY P10925 mus musculus (mouse).

zinc finger y-chromosomal prote 610
ZFY1 MOUSE MHQCRHCDFNSPDPFLLSHHlLSAIITKNV P10925 mus musculus (mouse).
zinc finger y-chromosomal prole 667
ZFY1 MOUSE VYQCEYCEYSTKDASGFKRHVISIHTKDY Pl0925 mus musculus (mouse).
zinc finger y-chromosomal prote 724
ZFY1 MOUSE PHSCDFCKKGFRRPSEKNQHIMRIIIIKVGL Pl0925 mus musculus (mouse).
zinc finger y-chromosomal prote 753
ZFY2_MOUSE TCKCKFCDYETAEQTLLNHHLLVVHRKKF P20062 mus musculus
(mouse). zinc finger y-chromosomal prote 496
ZFY2_MOUSE PYECQYCEYKSADSSNLKTHIKSKHSKEI P20062 mus musculus (mouse)-
zinc finger y-chromosomal prote 553
ZFY2_MOUSE THQCSI~CNHKSSNSSDLKRHIISVHTKAY P20062 mus musculus (mouse).
zinc finger y-chromosomal prote 610
ZFY2 MOUSE MHQCRHCDFNSPDPFLLSHHILSAIHTKNV P20062 mus musculus (mouse).
zinc finger y-chromosomal prote 667
ZFY2_MOUSE VYQCEYCEYSTKDASGFKRIIVISIHTKDY P20062 mus musculus (mouse).
zinc finger y-chromosomal prote 724
ZFY2_MOUSE PHRCDFCKKGFRRPSEKNQHIMRHHKEVG P20662 mus musculus
(mouse). zinc finger y-chromosomal prote 753
ZFY HUMAN MHKCKFCEYETAEQGLLNRHILLAVHSKNF P08008 homo sapiens
(human). zinc finger y-chromosomal prote 514
ZFY_HUMAN PYQCQYCEYRSADSSNLKTHIKTKHSKEM P08008 homo sapiens (human).
zinc finger y-chromosomal prote 571
ZFY HUMAN THQCLHCDHKSSNSSDLKRHVISVHTKDY P08008 homo sapiens (human).
zinc finger y-chromosomal prote 628
ZFY_HUMAN MHQCRHCDFKIADPFVLSRHILSVHTKDL P08008 homo sapiens (human).
zinc finger y-chromosomal prote 685
ZFY_HUMAN VYQCEYCEYSTTDASGFKRHVISIHTKDY P08008 homo sapiens (human).
zinc finger y-chromosomal prote 742
ZFY HUMAN PHRCEYCKKGFRRPSEKNQHIMRHHKEVG P08008 homo sapiens
(human). zinc finger y-chromosomal prote 771
ZG17_XENLA PFSCSECGKCF ARSSDLTVHRRRSHTKEK P18713 xenopus laevis (african
clawed frog). gastrula zinc f 145
ZG26 XENLA PFSCTECGKCFTIKSTLQSHLKRTHTGEK P18715 xenopus laevis (african
clawed frog). gastrula zinc f 285
ZG62 XENLA PFTCTDCGKCFSVKSILNHHRQAIHSGEK P18731 xenopus laevis (african
clawed frog). gastrula zinc f 89
ZG66 XENLA HDFCSECGKCFATSSQLIAHQQQVHIEVK P18733 xenopus laevis (african
clawed frog). gastrula zinc f 383
ZG66_XENLA PYSCSECGKCFASSSHLIGHRQQVHMEGK P18733 xenopus laevis (african
clawed frog). gastrula zinc f 440
ZG66 XENLA PYSCSECGKCFATSSQLMAHQQQVHIEVK P18733 xenopus laevis (african
clawed frog). gastrula zinc f 497
ZG66 XENLA PDFCFECGKCFATSLQLIAHQQQVHMEVK P 18733 xenopus laevis (african
clawed frog). gastrula zinc f 554
ZNFP_LYCVA PLSCKSCWQKFDSLVRCHDHYLCRHCLNL P18541 lymphocytic
choriomeningitis virus (strain Mmstrong) 29
ZNFP_LYCVP PLNCKSCWQKFDSLVRCHDHYLCRHCLNL P19326 lymphocytic
choriomeningitis virus (strain pasteur).
ZNFP_LYCVT PLNCKSCWQKFDSFSKCHDHYLCRHCLNL P19325 lymphocytic
choriomeningitis virus (strain traub). zi 29
Z0Z2 XENLA LHSCSQCGKCFSSSSDLLAHRRQSHTREK P18745 xenopus laevis (african
clawed frog). oocyte zinc fin 96
Z022 XENLA PYSCSECGKSFVTSSQLAVllRRRTHTGEK P18745 xenopus laevis (afriCarl
clawed frog). oocyte zinc fin 321
Z022 XENLA LFSCSECGKSFVTSSKLASHQRQTIITGEK P 18745 xenopus laevis (african
clawed frog). oocyte zinc fin 378
Z06 XENLA GFICSKCGETFTVNSHLLTIILCGKHERIY P18749 xenopus laevis (afriCarl
clawed frog). oocyte zinc fin 5
Z071 XENLA PFCCSECGKCFSSSSGLTAIHQQRTHMKVK P18751 xenopus laevis (african
clawed frog). oocyte zinc fin 537

Z071 XENLA PISCPECEECFVSSSQLTAHQQQAHRMVK P18751 xenopus laevis (afriCarl clawed frog). oocyte zinc fin 734

Προγράμματα που χρησιμοποιήθηκαν σ' αυτήν την εργασία

Πρόγραμμα για τον υπολογισμό των σχετικών συχνοτήτων (R.O.) και τυπικών αποκλίσεων (e.s.d's) για τα αντίστοιχα ζευγάρια.

```
PROGRAM compar
INTEGER resocl, resoc2, occsum, respsl, resps2, comocc
REAL sdf n, a, b, c
REAL respcl, respcZ, combpc, norpc
CJIARACTER*l postpl, postpZ, anposl, anposZ, restyp, anresl,
anres2
CHARACTER*23 aares
CHARACTER*80 filinl, filin2, cmpfil, outnl, outf1Z, inpstr
CHARACTER*122 outstr
DIMENSION respcl(20), respc2(20), combpc(20,20), comocc(20,20), sdf(20,20),
resocl(20), resoc2(20)
aares='KEQDNRSTPGAYHLMWVFIC'

WRITE(*, *)' Please enter datafile #1 :'
READ(*, '(A80)')filinl OPEN(11,ERR=999,FILE=filinl,STATUS='OLD')

WRITE(*, *)' Please enter datafile #2 .
READ(*,'(A80)')filinZ OPEN(12,ERR=999,FILE=filin2,STATUS='OLD')

WRITE(*, *)' Now enter the file to be analysed
READ(*, '(A80) ')cmpfil OPEN(14,ERR=999,FILE=mlpfi1,STATUS='OLD , ,
-       CARRIAGECONTR OL='LIST')


c Pass by headers in file 1

100 READ(11, '(A80)')inpstr
IF (inpstr(l:l).NE.'#') GO TO 100
READ(inpstr(8:8), '(Al) ')postpl

c Pass by headers in file 2
101     READ(12, '(A80)')inpstr
IF (inpstr(l.l).NE.'#')GO TO 101
READ(inpstr(8:8), '(Al)')postp2

c Pass by headers In file 3
102     READ(14, '(A80)')inpstr
IF (inpstr(l:l).NE.'#') GO TO 102 READ(inpstr(1:16),'(2(ZX,Al))')anposl, anpos2
IF ((Posrpl.NE.anposl).0R.(posrp2.NE.anpos2)) - STOP'***Positlon mismatch
In
the fi1es***'
WRITE(*,*) ' Doing the hard lob...'

c Read and normalise occurences in file 1
occsum=0
200     READ(11, '(A80)',END=201)inpstr
READ(inpstr(8:8), '(Al))restyp
```

```
respsl=INDEX(aares ,restyp) READ(blpstr(21:24), (I4)')resocl(respsl)
occsum=occsum+resocl(respsl)
GO TO 200
201     DO i=1,20
respcl(i)=FLOAT(resocl(i))/FLOAT(occsum)
ENDDO
c Read and normalise occurrences in file 2
occsum=0
210     READ(12, '(A8O)',END~211)inpstr
READ(inpstr(8:8), '(Al)')restyp
resps2=INDEX(aares,restyp)
READ(inpstr(21:24), '(14) ')resoc2(resps2)
occsum =occsum +resoc2(resps2)
GO TO 210
DO I=1,2O
respc2(i)=FLOAT(resoc2(i))/FLOA T(occsum)
ENDDO
occsum=0
 300              READ(14, '(A80) ',END~3O1)inpstr
READ(inpstr(1:16), '(2(7X,Al))')anresl,anres2
respsl=INDEX(aares,anresl)
resps2=INDEX(aares, anres2)
READ(inpstr(21:24), '(I4)')comocc(respsl,resps2)
occsum~occsum+comocc(respsl,resps2)
GOTO 300
301     DO i1=1,20
 DO i2=1,20
 n=FLOAT(occsum)
 c=FLOAT(comocc(il,i2))
 a=FLOAT(resocl(il))
 b=FLOAT(resoc2(i2))
 combpc(il,iZ)=c/n
IF ((a.NE.O.O).AND.(b.NE.O.O)) THEN
sdf(il,i2)=(n*n)*( (c/(b*a))*(c/(b*a*a)) +
 + (c/(a*b))*(c/(a*b*b)) + + (c/(a*b))*(l/(a*b)) )
 sdf(il,i2)=SQRT(sdf(il,i2))
ELSE
sdf(il,i2)=-1.0
ENDIF
ENDDO
ENDDO
WRITE(*, *) ' ..........................'
WRITE(*, *)' That was easy.'
WRITE(*, *)' Output file for number of times:'
READ(*, '(A80)')outfll
OPEN(15,ERR=998,FILE=outfll,STATUS=WEW'
CARRIAGECONTROL='LIST')
WRITE(15, *)' Relative occurrence'
WRITE(*, *)' Output file for SDs .
READ(*, '(A80) ')outfl2 OPEN(16,ERR=99S,FILE=outfl2,STATUS=NEW'
CARRIAGECONTROL='LIST'),
WRITE(16, *)' SD of Relative occurrence'
```

```
DO il=1,122
        outstr(il:il)=’   ‘,
ENDDO
DO il=1,20
        outstr(il*6+2:il*6+2)=aares(il:il)
ENDDO
WRITE(15, '(A122)')outstr
WRITE(16, '(A122)')outstr

DO il=1,20
        DO i2=1,122
         outstr(i2:i2)=’ ‘ ,
ENDDO
outstr(2:2)=aares(il:il)
DO i2=1,20
  IF ((respcl(il).NE.0.0).AND.(respc2(i2).NE.0.0)) THEN
        norpc=ombpc(il,i2)/(respcl(il)*respc2(i2))
        WRITE(outstr(i2*6-3:i2*6+2), '(F6.2)')norpc
ELSE
        WRITE(outstr(i2*6-3:i2*6+2), '(A6)')' ---~
        ENDIF
ENDDO
WRITE(15 '(A122)')outstr
DO i3=3,122 outstr(i3:i3)=’  ‘ ,
ENDDO
DO i2=1,20
IF (sdf(il,i2).NE.-1.0) THEN
WRITE(outstr(i2*6-3:i2*6+2), '(F62)')sdf(il,i2)
ELSE
        WRITE(outstr(i2*6-3:i2~6+2), '(A6)')' ----‘
        ENDIF
ENDDO
WRITE(16 '(A122)')outstr
ENDDO

CLOSE(11)
CLOSE(12)
CLOSE(14)
CLOSE(15)
CLOSE(16)

STOP'Ok'
998     STOP'***Error creating output file***'
999 STOP'***Error accessiog ioput file***' END
```

Πρόγραμμα για τον υπολογισμό των αντιστοίχων ποσοτήτων για τις τριπλέτες

```
PROGRAM compar3
INTEGER     resocl, resoc2, resoc3, occsum, respsl, resps2,
```

```
      resps3, comocc, ooozer , monii, moonj, monkk

REAL        sdf, n, 8, b, c, d
REAL        respcl, respc2, respc3, combpc, norpc, monpin
CHARACTER*1 postpl, postpZ, postp3, anposl, anposZ, anpos3, - restyp,
anresl, anresZ,
8nres3 CHARACTER*20 8ares
CHARACTER*80 filinl, filin2, filin3, cmpfil, outfl1, inpstr
CHARACTER*24 outstr
DIMENSION respcl(20), respc2(20), respc3(20), combpc(20,20,20),
-       comocc(20,20,20), sdr(20,20,20), resocl(20),
-       resoc2(20), resoc3(20), monpin(8000), monii(8000),
monjj(8000),monkk(8000)
aares='KEQDNRSTPGAYHLMWVFIC'

WRITE(*,*)' Please enter d8tafile #1 :,
READ(*, '(A80) ')filinl
OPEN(11,ERR=999,FILE=filin1,STATUS='OLD')
WRITE(*,*)' Ple8se enter d8tafile #2 :,
READ(*, '(A80)')filin2
OPEN(12,ERR=999,FILE=fi1in2,STATUS='OLD')
WRITE(*,*)' Ple8Se enter d8tafile #3 ;,
READ(*, '(A80)')filin3
OPEN(13,ERR =999,FILE=fi1in3,STATUS='OLD')
WRITE(*,*)' Now enter the file to be ana1ysed .
READ(*, '(A80)')cmpfil
OPEN(14,ERR=999,FILE=cmpfH,STATUS='OLD'
- CARRIAGEC0NTROL='LIST') ,
c Pass by headers in file 1
101     READ(11, '(A80)')inpstr
IF (inpstr(l:l).NE.'#') GO TO 101
READ(inpsfr(8:8), '(Al)')postpl
c Pass by headers in file 2
102     READ(12, '(A80)')inpstr
IF (inpsfr(l:l).NE.'#')
GO TO 102 READ(inpstr(8:8), '(Al)')postp2

c Pass by headers in file 3
103     READ(13, '(A80) ')inpstr
IF (inpstr(l:l).NE.'#')
GO TO 103 READ(inpstr(8:8), '(Al)')postp3

c Pass by headers in comp. file
104     READ(14, '(A80) ')lnpstr
IF (inpsfr(l:l).NE.'#') GO TO 104
READ(inpsfr(1:24),'(3(7X,Al))')anposl, anpos2, anpos3
IF((Postpl.NE.anposl).OR.(postp2.NE.anpos2).OR.(postp3.NE.anpos3)) –
STOP'***Position mismatch in the files***'

WRITE(*,*)' Doing the hard job...''

c Read and nonn81ise occurrences in file 1 occsum=o
```

```
210     READ(11.'(A80)' END=211)inpstr
READ(inpsrr(8:8), '(Al)')resryp
respsl=INDEX(aares,resryp)
READ(inpsrr(21:24), '(I4)')resocl(respsl)
occsum=occsum+resocl(respsl)
GO TO 210
DO i=1,20 respcl(i)=FLOAT(resocl(i))/FLOA T(occsum)
ENDDO
  c   Read and normalise occurrences in file 2
occsum=0
READ(12, '(A80)',END=221)inpstr
READ(inpstr(8:8), '(Al)')resryp
resps2=INDEX(aares,restyp)
READ(inpsrr(21:24), '(I4)')resoc2(resps2)
occsum =occsum +resoc2(resps2)
GO TO 220
221     DO i=1,20
respc2(i)=FLOAT(resocZ(i))/FLOAT(occsum)
ENDDO
     c  Read and normalise occurrences in file 3
occsum=0
READ(13, '(A80)',END=231)inpsrr
READ(inpstr(8:8), '(Al) ')resryp
resps3=INDEX(aares,resryp)
READ(inpstr(21:24), '(I4)')resoc3(resps3)
occsum =occsum +resoc3 (res ps3)
GO TO 230
231     DO i=1,20
respc3(i)=FLOAT(resoc3(i))/FLOA T(occsum)
ENDDO
occsum=0
300     READ(14, '(A80)',END=301)inpstr
READ(inpstr(1:24), '(3(?X,Al)) ')anresl,anres2,anres3
respsl=INDEX(aares,anresl)
resps2=INDEX(aares, anres2)
resps3=INDEX(aares, anres3)
READ(inpstr(29:32), '(14) ')comocc(respsl,resps2,resps3)
Occsum=occsum +comocc(respsl,resps2,resps3)
GOTO 300
301     DO il=1,20
  DO i2=1,20
             DO i3=1,20
      n=FLOAT(OCCSUIO)
                d=FLOAT(comocc(i1,i2,i3))
a=FLOAT(resocl(il))
b=FLOAT(resoc2(i2))
c=FLOAT(resoc3(i3))
combpc(il,i2,i3)=d/n
IF ((a.NE.0.0).AND.(b.NH.0.0).AND.(c.NH.0.0)) THEN
   sdf(il,i2,i3)=(n**4)*( (d/(b*a*c))*(d/(b*a*a*c)) +
+                 (d/(a*b*c))*(d/(a*b*b*c)) +
+                 (d/(a*b*c))*(d/(a*b*c*c)) +
```

```
+                    (d/(a*b*c))*( l/(a*b*c))
sd(il,i2,i3)=SQRT(sdf(il,i2,i3))
ELSE
sdf(il,i2,i3)= -1.0
  ENDIF
  ENDDO
 ENDDO
ENDDO

C        Sort out numbers
non.zer=l
DO il=1,20
  DO i2=1,20
    DO i3=1,20
      IF ((respcl(il).NE.0.0).AND.(respc2(i2).NE.0.0)
–                  .AND. (respc3(i3).NE.0.0)) THEN
norpc=combpc(il,i2,i3)/(respcl(il)*respc2(i2)*respc3(i3))
monpin(non.zer)=norpc
monii(non.zer)=il
monjj(non.zer)=i2
monkk (non.zer) =i3
if (norpc.NE.0.0) non.zer=non.zer+l
    ENDIF
  ENDDO
 ENDDO
ENDDO
CALL sort(monpin,monii,monjj,monkk,anpos3,nonzer)
WRITE(*,*)' ..... '
WRITE(*,*)' That was easy.'
WRITE(*,*)' Output file :'
READ(*, '(A80)')outfll
OPEN(15,ERR=998,FILE=outfll,STATUS='NEW'
CARRIAGECONTROL='LIST') ,
WRITE(15, '(3(lX,Al,lX),A21)')anposl,anpos2,anpos3,
'R-occ SD RO/SD'

DO il=l,nonzer
IF(monpin(il).NE.0.0) THEN
IF(sdf(monii(il),monjj(il),monkk(il)).NE.0.0) THEN
WRITE(15,'(3(lX,Al,lX),3F7.2)')
aares(monii(il):monii(il)),aares(monjj(il):monjj(il)),
aares(monkk(il):monkk(il)),monpin(il),
sdf(monii(ll),monjj(il),monkk(il)),
monpin(il)/sdf(monii(il),monjj(ll),monkk(il))
        ELSE
WRITE(15, '(3(lX,Al,lX),2F7.2,3X,A4)')
aares(monii(il):monii(ll)),aares(monjj(il):monjj(il)),
aares(monkk(ll):monkk(ll)),monpin(ll),
sdf(monii(ll),monjj(il),monkk(il)), '----'
ENDIF
ENDIF
ENDDO
```

```
CLOSE(11)
CLOSE(12)
CLOSE(14)
CLOSE(15)

STOP'Ok'
998     STOP'***Error creating output file***'
999     STOP'***Error accessing input file***'
END

SUBROUTLNE sort(a,ii,jj,kk,gramma3,n}

     CHARACTER*l gramma3
INTEGER bound,last,i,n,tempii,tempjj,tempkk
INTEGER ii, jj, kk
    REAL a, temp
    DIMENSION a(*), ii(*), jj(*}, kk(*)
Bound=n
1         last=0
do 10 i=l,bound-l
if (a(i).lt.a(i+l)) then
temp=a(i)
tempii=ii(i)
tempii=ii(i)
if(gramma3.ne.' ') tempkk=kk(i)

a(i)=a(i+l)
ii(i)=ii(i+l)
jj(i)=ii(i+l)
if(gramma3.ne.‘ ’) kk(i)=kk(i+l)
a(i+l)=temp
ii(i+l) =tempii
jj(i+l)=tempjj
if(gramma3.ne.’ ’) kk(i+l)=tempkk
   last=i
endif

10       continue
            bound=last
if (bound.ne.0) goto  1

return
end
```

Πρόγραμμα για τον υπολογισμό των συχνοτήτων ενός, ή παραπάνω αμινοξέων στις αντίστοιχες θέσεις

```
PROGRAM taxinomhsh
INTEGER numpos, linpos, regstr , rawdat,str, thesh, vas
```

```
INTEGER      maxthesh,maxstr,secraw,rege,regl,numreg,pin2
INTEGER      gnuml,gnum2,gnum3,pinak,rgstl,rgst2,rgst3
INTEGER      monpin,monii,monjj,monkk,m
CHARACTER*1 posryp,restyp,r,resl
CHARACTER*1  rsrpl,rsrp2,rsrp3,grammal,gramma2,gramma3
CHARACTER*23 aares
CHARACTER*36 oursrr
CHARACTER*80 inpstr, inpfil, ourfil,ourfiIl
DIMENSION posryp(100),linpos(1W),rawdar(23,1W),pin2(24,24)
DIMENSION man(23,100),srr(23,1W),rhesh(23,100), vas(lOO)
DIMENSION secraw(23,100,100),rege(23,l00),pinak(24,24,24)
DIMENSION monpin(l000),monii(l000),monjj(1000),monkk(1000)
aares='KEQDNRPSTGAYHMLWVFICBZX'
t=CHAR (9)

WRITE(*,*)'Please enter the filename:'
READ(*, '(A80) ')inpfil
write(*,*) ' '

OPEN(11,ERR=999,FILE=inpfil,STATUS='OLD'
c         1 CARRIAGECONTROL='LIST')

100     READ(ll, '(A80)')inpsrr
IF (inpsrr(l:l).NE.'#') GO TO 100
numpos=0
DO lcntl=2,80
 IF (inpstr(lcnrl:1cnrl).EQ.' GOTO  199
numpos=numpos+l
postyp(numpos)=blpstr(lcnrl:1Q2rl)
lfup 05 (n umpos ) =lcn t 1
199     CONTINUE
ENDDO
IF (numpos.EQ.O) STOP'***No posidon types read***'
WRITE(*, *) 'Number of positions =',numpos
write(*,'(Al)') , CHARACTERS . '' (postyp(i),i=l,numpos)
wrire(*, *) , ,
wrire(*,*) 'PLEASE ENTER 3 CHARACTERS (e.x. a,b,c)'
C
wrire(*, *) 'character number 1 .....?????'
read(*,'(Al)') gramma1
write(*, *) 'character number 2 .....?????'
read(*,'(Al)') gramma2
wtite(*,*) 'IF YOU WANT ONLY 2 CHARACTERS PRESS RETURN .....!!!!
write(*, *) 'chartacter number 3 ....?????'
read(*,'(Al)') gramma3
do 1199 i=l,numpos
if(postyp(i).eq.grammal) gnuml=i
if(posryp(i).eq.grammaZ) gnum2=i
if(gramma3.eq.' ') goto 1199
if(postyp(i).eq.gramma3) gnum3=i
1199   continue
write(*, *) 'POSITIONS OF CHARS : ',gnuml,gnum2,gnum3
```

```
2001   READ(11, '(A80)', END=8900)inpstr
rstpl=inpstr(linpos(gnuml):linpos(gnuml))
rgstl=INDEX(aares,rstpl)
rstp2=inpstr(Ihlpos(gnum2):linpos(gnum2))
rgst2=INDEX (aares,rstp2)
if (gramma3.eq.' ') then
  pin2(rgstl,rgst2)=1+pin2(rgstl,rgst2)
  goto 8000
endif
rstp3=inpstr(linpos(gnum3):linpos(gnum3))
rgst3=LNDEX(aares,rstp3)
pinak(rgstl,rgst2,rgst3)=1+pinak(rgstl,rgst2,rgst3)
8000   continue
GO TO 2001
8900   continue

       if (gramma3.eq.' ') goto 6012
C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C

WRITE(*,*)' Please enter Jilename (1) for output:'
READ(*, '(A80) ')outfill
OPEN(25,ERR=998,FILE=outfill,STATUS='UNKNOWN#)
c      write(*,*) , grammal
m=l
do 7800 i=1,23
do 7801 j=1,23
do 7802 k=1,23
m onpin (m) =pinak (i,j, k)
monii(m)=i monjj(m)=j
monkk(m)=k
if (pinak(i,},k).eq.O) goto 7802
write(*,*) i,j,k,' ',pinak(i,j,k)
m=m+l
7802   continue
7801   continue
7800   continue

call sort(monpin,monii,monjj,monkk,gramma3,m)
write(25, *) '*************** TAXINOMHMENOS PINAKAS
*************'

write(25, '(3(7X,Al),5X,A3)')grammal,gramma2,gramma3, 'sum'
do 7888 i=l,m
if(monpin(i).eq.0) goto 7888
write(25, '(3(7X,Al),I8) ')
aares(monii(i):monii(i)),aares(monjj(i):monjj(i)),
aares(monkk(i):monkk(i)),monpin(i)
7888   continue
close (25)
```

```
goto 6013
C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
6012    WRITE(*,*)' Please enter filename (1) Jor output:
READ(*, '(A80)')outfill
OPEN(25,ERR=998,FILE=outfi1l,STATUS='UNKNOWN')
C       write(*,*) , grammal gramma2       number'
m=l
do 7821 i=1,23
do 7822 j=1,23
monpin(m)=pin2(i,j)
monii(m)=i monjj(m)=j
monkk(m)=o
if (pinZ(i,j).eq.0) goto 7822
c       write(*,*) i,j,' ',pin2(i,j)
m=m+l
7822    continue
7821    continue

call sort(monpin,monii,monjj,monkk,granHOa3,m)

write(25,*) '**************88 TAXINOMHMENOS PINAKAS
***************'

write(ZS, '(2(7X,Al),5X,A3)')grammal,gramma2, 'sum ,
do 7898 i=l,m
if(monpin(i).eq.O) goto 7898
write(25 '(2(7X,Al),I8)') –
aares(monii(i) :monii(i)),aares(monjj(i) :monjj(i)),monpin(i)
7898 continue close (25)

6013    continue
C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
890     CLOSE(11)
goto 1000

998     STOP'~**EITor accessiog output file***'
999     999 STOP'***File oot fouod***'
1000    STOP' *** SUCCESS ---- THE FILE IS READY ***,
END

CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

 SUBROUTINE sort(a,ii,jj,kk,gramma3,n)
CHARACTER*1 gramma3
```

```
INTEGER bound,la.~t,i.n,tempii,tempjj.tempkk
INTEGER a(lOW),ii(lOOO),jj(lOOO),kk(lOW),temp

bound=n
C          write(*,*) 'BOUND' bound
last=0
do 10 i=l,bound-l

if (a(i).lt~a(i+l)) then
temp=a(i)
  tempii=ii(i)
  tempjj=jj(i)
  if(gramma3.ne.' ') tempkk=kk(i)

a(i)=a(i+l)
  ii(i)=ii(i+1)
  jj(i)=jj(i+l)
  if(gramma3.ne.' ') kk(i)=kk(i+l)

a(i+l)=temp
  8(i+1)=temp8
  ii(i+l)=tempii
  if(gramma3.ne.' ') kk(i+l)=tempkk

last=i
endif

10      contblue
bound=last
if (bound.ne. 0) goto 1

return
end
```

Πρόγραμμα για την δημιουργία της βάσης δεδομένων δακτύλων ψευδαργύρου

```
#include <stdio.h>
#include <stdlib.h>
#define BUFFSIZE 65535
#define CODESIZE 12
#define DESCRWID 60

char aacid[24]="ABCDEFGHIKLMNPQRSTVWXYZ".
main()
{
   int cnt0, cntf, cnt2, cnt3, cnt4, scntf, scnt2, wh;
   int ofs f, ofs2, ofs3, flgx;
   int zinc fou, fing_fou;
   int indx();
```

```
  char namein[6]="swiss", nameout[9]="zinc.out".
  char sbuf[BUFFSIZE] code[CODESIZE],

descripf[DESCRWID];
FILE "in , *out,;
inf chl='\n', ch2;

if ((in=fopen(namein,"r")) == NULL)
[prinff("%s not found \n" namein); exit(l);}

if ((out=fopen(nameout,"w')) == NULL)
[printf("Can not create %s \n",nameout); exit(l);}

flgx=0;
while ((ch2=getc(in)) EOF {
/* 15 it code (>>>>) line ? */
if ((chl == '\n' && (ch2 == '>')) {

/* If yes readln three lines */
buf[(cntl=1)]=ch2;
 while ((ch2=getc(in)) !='\n'
if ((++cntl) > BUFFSIZE)
    {printf("BUFFSIZE limit exceeded \n"}; exit(l};}
    buf[cntl]=ch2;
}/*endwhile*/
buf[cntl+l]='\n';

/* Parse it and find code */
cnt2=1;
while (buf[cnt2] == '>' cnt2++;  /* Pass by > signs */
if (buf[cnt2] == '') ,
[while (buf[cnt2] == '') cnt2++ ' } /* Pass by any possible spaces */
cnt3=0;
while ((buf[cnt2] !='') && (buf[cnt2] != '\n') &&
(buf[cnt2] != 0 ) && (cnt3 < CODESIZE))
[code[++cnt3]=buf[cnt2]; cnt2++;}
while (cnt3 < CODESIZE) code[++cnt3]='';

/* Read the protein description line */
cntl=0;
while ((ch2=getc(in)) != '\n')
[ if ((++cntl) > BUFFSIZE)
{printf("BUFFSIZE" limit exceeded "\n"); exit(l);}
buf[cntl]=ch2;
/*endwhile */

/* Read the sequence line */
scntl=0;
while ((ch2=getc(in}) != '>' ; {
if(ch2 == EOF)flgx=1;
if ((wh=indx(ch2))!= 24) {
if ((++cnt1)>BUFFSIZE)
```

```
{printf"BUFFSIZE" limit exceeded \n");  exit(l);}
subf[scnt1]=ch2;
}*endif*/
}/*endwhile */
if (flgx == 0) {
ch1='\n';
lungetc(ch2,in);

ungetc(ch1,in);}
else {
chl=EOF;
 ungetc(ch1,in);}

/* Try to find 'zinc' and 'fing' in the name */
zinc_fou=0;
fing_fou=0;
for (cnt2=1; cnt2<=cntl-3; cnt2++) {
    if ( ((buf[cnt2] == 'z') || (buf[cnt2] =='Z')) &&
((buf[cnt2+ 1] == 'i') || (buf[cnt2+11 =='I')) &&
((buf[cnt2+2] == 'n') || (buf[cnt2+21 =='N')) &&
((buf[cnt2+3] == 'c') || (buf[cnt2+31 =='C')) ) zinc_fou=l;

if ( ((buf[cnt2] == 'f') || (buf[cnt2] =='Z')) &&
((buf[cnt2+ 1] == 'i') || (buf[cnt2+1] =='I')) &&
((buf[cnt2+2] == 'n') || (buf[cnt2+2] =='N')) &&
((buf[cnt2+3] == 'g') || (buf[cnt2+3] =='C')) ) zinc_fou=l;
}/*endfor*/


if ((zinc_fou == 1) && (fing_fou == 1)) {

for (scnt2=l; scnt2<=scntl; scnt2++) {
 if (sbuf[scnt21 == 'C') {
  for (ofsl=3; ofsl<=5; ofsl++) {
    if (scnt2+ofs1 > scntt) goto out_of_ran;
      if (sbuf[scnt2+ofsl] == 'C') {
       for (ofs2=12 ofs2<=14; ofs2++) {
        if (scnt2+ofs1+ofs2 > scnt1) goto out_of_ran;
          if (sbuf[scnt2+ofs1+ofs2] == 'H' {
             for (ofs3=4; ofs3<=5; ofs3++) {
              if (scnt2+ofs1+ofs2+ofs3 > scnt1) goto out_of_ran;
                  if (sbuf[scnt2+ofs1+ofs2+ofs3] == 'H') {

cnt2=t;
while ((buf[cnt2] != 0) && (buf[cnt2] != '\n') && [while (cnt2 <= DESCRWID)
{descript[cnt2]=buf[cnt2]; cnt2++;}
while(cnt2 <= DESCRWID) descript[cnt2++1]=' ';

for (cnt4= 1; cnt4<=CODESIZE; cnt4++) putc(code[cnt4],stdout);
for (cnt4=1; cnt4<=DESCRWID; cnt4++) putc(descript[cnt4] stdout);
fprintf(stdout" %d \n",scnt2);
```

```
for (cnt4= 1; cnt4<=CODESIZE; cnt4++) putc(code[cnt4],out);
putc(' ',out);
for (cnt4=0-1; cnt4<=ofs 1+ofs2+ofs3+2; cnt4++)
     putc(sbuf[scnt2+cnt4-1],out);
putc(' ',out); putc(' ',out);
for (cnt4=1; cnt4<=DESCRWID; cnt4++) putc(descript[cnt4],out);
fprintf(out," %d \n",scnt2);

 }/*endif*/
}/*endfor*/

            }/*endif*/
           }/*endfor*/
         }/*endif*/
        }/*endfor*/
       }/*endif*/
      }/*endfor*/
      }/*endif*/

out_of_ran: chl='\n';

}/*endif*/
else
 chl=ch2;

}/*endwhile*/
 fclose(in);
fclose(out);
}

int indx(ch)
 int ch;

int cl, post=24;
for (cl=0;  cl<=22; cl++} { if (aacid[cl]==ch) return cl;}
return post;
}
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

Altaba A.R. et al., (1988) " Xfin: an embryonic gene encoding a multifingered protein in Xenopus" EMBO J **6**, 3065-3070.

Berg J.N. (1988) "Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins" Proc.Natl.Acad.Sci.USA **85**, 99-102.

Berg J.N., van Opheusden J.H.J., Burgering M.J.M., Boelens R., Kaptein R. (1990) "Structure of Arc repressor in solution: evidence for a family of â-sheet DNA-binding proteins" Nature **346**, 586-589.

Brennan R.G., Roderick S.L., Takeda y. and Matthews B.W. (1990) "Protein DNA conformation changes in the crystal structure of a 1ambda cro-operator complex" Proc.Natl.Acad.Sci.USA **87**, 8165-8169.

Dejarlais J.R., Berg J.M. (1992) "Toward rules relating zinc finger protein sequences and DNA binding site preferences" Proc.Natl.Acad.Sci.USA **89**, 7345-7349.

Desjarlais J.R., Berg J.M. (1992) "Redesigning the DNA-Binding of a Zinc Finger Protein: A Data Base-Guided Approach" Proteins **12**,101-104.

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

Gibson T.J., Postma J.P.M., Brown R.S. and Argos P. (1988) "A model

for the tertiary structure of the 28 residue DNA-binding motif , zinc finger' common to many eukeryotic transcriptional regulatory proteins" Protein Engineering **2**, 209-218.

Gogos J.A., Hsu T., Bolton J., Kafatos F.C. (1992) "Sequence Discrimination by Alternatively Spliced Isofonns of a DNA Binding Zinc Finger Domain" SCIENCE **257**, 1951-1954

Hard T., Dahlman D., Carlstedt-Duke J., Gustafsson J.A. and Rigler R. (1990) "Cooperativity and specificity in the interactions between DNA and the glycocorticoid receptor DNA-binding domain" Biochemistry **29**, 5358-5364.

Harris L.F., Sullivan M.R., Hickok D.F. (1992) "Conservation of genetic infonnation: A code for site-specific DNA recognition" Proc.Natl.Acad.Sci.USA **90**, 5534-5538.

Harrison S.C. (1991) "A structural taxonomy of DNA-binding domains" Nature **353**, 715-719.

Hsu T., Gogos J.A., Kirsh S.A., Kafatos F.C. (1992) "Multiple Zinc Finger Forms Resulting from Developmentally Regulated Alternative Splicing of a Transcription Factor Gene" SCIENCE **257**, 1946-1950.

Jacobs G.H. (1992) "Determinetion of the base recognition positions of zinc fingers from sequence analysis" EMBO J **11**, 4507-4517

Jaenicke R. (1987) Pergamon (ed.) "FOLDING AND ASSOCIATION OF PROTEINS" Institut fur Biophysik and Physikalische Biochemie der Universitat Regensburg, 117-237.

Jones T.A. (1988) In Sayre,D. (ed.),"Computational Crystallography" Chaledron Press, Oxford, 303-317.

Jones T.A., Zou J.-Y., Cowan S.W. (1991) "Improved methods for binding protein models in electron density maps and the location of errors in these models" Acta Cryst. **47**, 110-119.

Jordan S.R., Pabo C.O. (1988) "Structure of the 1ambda repressor at 2.5 A resolution: details of repressor-operator interactions" Science **242**, 893-899.

Kaptein R., Zuiderweg E.R.P., Scheek R.M., Boelens R. and van Gunsteren W.F., (1985) "A protein structure from nuclear magnetic resonanse data; lac repressor headpiece" J. Mol. Blo. **182**, 179-182.

Kissinger C.R., Liu B., Martin-Blanco E., Korberg T.B. and Pabo C.O. (1990) "Crystal Structure of an engralled Homeodomaln-DNA complex at 2.8 A Resolution: A Framework for understanding Homeodomain-DNA Interactions" Cell **63**, 579-590.

Klevit R.E., Herrlot J.R. and Horvath S.J. (1990) "Solution structure of a zinc finger domain of yeast ADRl" Proteins **7**, 215-226.

Kostrewa D., Granzin J., Koch C., Choe H.W., Raghunathan S., Wolf W., Labahn J., Kahmann R. and Saenger W. (1991) "Three dimensional structure of the E.coli DNA-binding protein Fis" Nature **349**, 178-180.

Landschultz W.H., Johnson P.F. and McKnight S.L. (1989) "The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite" SCIENCE **243**, 1681-1688.

Lifson S. (1972) R.Jaenicke and B.Helreich, (ed.) "Molecular Forces in Protein Protein Interactions" CoUoquium Mosbach **33**, 3- 16.

Lee M.S., Gippert G.P., Soman K.V., Case D., Wright P.E. (1989) "Three-Dimensional Solution Structure of a Single Zinc Finger DNA- Binding Domain" SCIENCE **245**, 635-637

Lifson S. (1973) B.PuUman (ed.) "In Conformation of Molecules and Polymers" The Jerasulem Symp. Quantum Chemistry and Biochemistry 5.

Lifson S. (1980) M.Balaban (ed.) "In Molecular Structure and Dynamics"

Int. Sci. Services Philadelphia, 213-243.

Luisi B.F. and Sigler P.B. (1990) "The stereochemistry and biochemistry of the trp repressor-operator complex" Biochim. Biophys. Acta **1048**, 113-126.

Mondragon A., Wolberger C. and Harrison S.C. (1989) "Structure of phage 434 cro protein at 2.35 A resolution" J Mol. Bio. **205**, 179-188

Mondragon A., Subbiah S., Almo S.C., Drottar M. and Harrison S.C. (1989) "Stmcture of the arnino-terminal domain of phage 434 repressor at 2.0 A resolution" J. Mol. Bio. **205**, 189-200.

Nardelli J., Gibson T.J., Vesque C., Charndy P. (1991) "Base sequence discrimination by zinc-finger DNA-binding domains" LETTERS TO NATURE **349**, 175-178.

Novotny J., Bruccoleri R.E. and Karplus M. (1984) "Analysis of incorrectly folded protein models. Implications for structure predictions" Mol.Biol **177**, 787791.

Oas T.G., Mclntosh L.P., O'Shea E.K., Dahlquist F.W. and Kim P.S. (1990) "Secondary structure of a leucine zipper determined by nuclear magnetic spectroscopy" Biochemistry **29**, 2891-2894.

Otwinowski Z., Schewitz R.W., Zhang R.G., Lawson C.L., Joachimiak A., Marmostein R.Q., Luisi B.F. and Sieger P.B. (1988) "Crystal structure of trp repressor/operator complex at atomic resolution" Narure **335**, 321-329.

Overington J., Johnson M.S., Sali A., Blundell T.L. (1990) "Tertiary structural constraints on protein evolutionary diversity. templates, key residues and structure predictions" Proc.R.Soc.Lond.B **241**, 132-145.

Paliakasis C.D., Kokkinidis M. (1992) "Relationships between sequence and structure for the four-a-helix bundle tertiary motif in proteins" Protein Engineering **8**, 739-748.

Pavletich N.P., Pabo C.O. (1991) "Zinc finger-DNA recognition: Crystal structure of a Zif268~DNA complex at 2.1 A" SCIENCE **252**, 809-816.

Philips S.E.V., Manfield I., Parsons I., Davison D.E., Rafferty J.B., Somers W.S., Magarita D., Cohen G.N. Saint-Girons I. And Stockley P.G. (1989) “Cooperative tandem binding of met repressor of Escherichia coli” Nature **341**, 711-715.

Puglisi J.D., Chen L., Frankll A.D., Williamson J.R. (1993) "Role of RNA arginine recognition of TAR RNA" Proc.Natl.Acad.Sci. USA **90**, 3680-3684.

Qian Y.Q., Billeter M., Otting G., Muller M., Gehring W.J. and Wuthrich K., (1989) "The structure of the Antennapedia homeodomain detennined by NMR spectroscopy in solution: comparison with prokaryotic repressors" Cell **59**, 573-580.

Rafferty J.B., Somers W.S., Saint-Girons 1. and Philips S.E.V. (1989) "Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor" Nature **341**, 705-710.

Rhodes D. and Klug A. (1986) "An underlying repeat in some transcriptional control sequences corresponding to half a double helical turn of DNA" Cell 46, l23~132. Richardson J~S. and Richardson D.C. (1989) In Fasman G., (e d.) "Prediction of Protein Structure and the Principles of Protein Conformation" Plenum Press, New York, 1-98.

Richardson J.S. and Richardson D.C. (1988) "Amino Acid Preferences for Specific Locations at the Ends of the a Helices" SCIENCE 240, 1648-1652.

Saenger W. (1984) In Saenger W., (ed.) "Principles of Nucleic Acid Structure" Springer-Verlag Inc.

Schultz S.C., Shields G.C. and Steitz T.A. (1990) "Crystallization of

Escherichia coli catabolite gene activator protein with its DNA binding site. The use of modular DNA" J Mol. Biol. Bio. **213**, 159-166.

Schwabe J.W., Neuhaus D. and Rodes D. (1990) "Solution structure of the DNA-binding domain of the oestrogen receptor" Nature **348**, 458-461.

Shimizu N., Ohta M., Fujiwara C., Sagara J., Mochizuki N., Oda T., Utiyarna H. (1992) "A Gene coding for a Zinc Finger Protein Is Induced during 12-0-Tetradecanoylphorbol- 13-Acetate-Stimulated HL-60 Cell Differentiation" J.Biochemistry **111**, 272-277.

White S.W., Appelt K., Whilson K.S. and Tanaka 1. (1988) "Structural comparison of the prokaryoticribosomal proteins L7/L12 and L30" Proteins 5, 281-288.

Yoon C., Prive G.G., Goodsell D.S., Dickerson R.E. (1988) "Structure of an altemating-B DNA heUx and its relationship to A-tract DNA" Proc.Natl.Acad.Sci.USA **85**, 6332-6336.