

Title	NMR Studies of the DNA-and RPA-Binding Domain of Human DNA Repair Protein XPA
Author(s)	池上, 貴久
Citation	大阪大学, 1999, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3155667
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

NMR Studies
of the DNA- and RPA-Binding Domain
of Human DNA Repair Protein XPA

(ヒト遺伝子修復蛋白質 XPA の DNA および RPA 結合ドメインの NMR による研究)

The doctoral thesis submitted to
Osaka University
1999

Takahisa Ikegami
池上 貴久

論文内容の要旨

NMR Studies of the DNA- and RPA-Binding Domain of Human DNA Repair Protein XPA (ヒト遺伝子修復蛋白質 XPA の DNA および RPA 結合ドメインの NMR による研究)

遺伝子修復系は、放射線、紫外線、化学変異剤、複製エラーなどで損傷を受けた DNA を修復する、生体維持に必須な機構である。とりわけヌクレオチド除去修復 (NER) は、多種多様な DNA 損傷に対処でき、真核生物では約 30 種類の蛋白質が関与すると考えられている。当研究の対象である XPA 蛋白質は、真核生物の一連の NER 機能の内、初期段階である塩基損傷部位の認識と修復系蛋白質の複合体形成の一部を担うと考えられている。XPA を含めいくつかの NER 因子の変異は、常染色体劣性遺伝病である色素性乾皮症 (Xeroderma Pigmentosum) の原因である。XP 患者は、紫外線などで損傷を受けた DNA の修復機構に障害を持つため、高頻度の皮膚癌、臓器癌、神経変性、知能障害などの症状をおこす。

全長 273 残基からなるヒト XPA 蛋白質が認識する DNA 損傷の種類は、紫外線による損傷から cisplatin などの化学物質による損傷までの広い範囲にわたる。Zinc-finger 領域を含む中央ドメイン (Met98-Phe219, 122 残基) は、損傷 DNA に選択的に結合し、また遺伝子複製にも関与する蛋白質 RPA (replication protein A) との結合能を有する最小ドメインであることが分かっている。筆者は、XPA 蛋白質が多種類の DNA 損傷を認識する機構と他の蛋白質と相互作用する機構を立体構造の観点から研究するため、核磁気共鳴法 (NMR) を用いて、XPA 中央ドメインの立体構造の決定、運動性の解析、および DNA、RPA70 と相互作用する部位の決定を行った。

まず、 ^{15}N , ^{13}C , (80% ^2H) で安定同位体ラベルした中央ドメインを精製し、一連の多核共鳴多次元 NMR 測定により、各原子の化学シフトの帰属、 ^1H 間距離情報である NOE スペクトルの解析、二面角度情報であるスカラー結合定数の測定、水素結合情報であるアミド水素/重水素交換速度の測定を行った。1,389 個の距離制限 (水素結合制限 47 個を含む) および、83 個の角度制限をもとに、simulated annealing 法により立体構造を計算した。その結果、中央ドメインは、亜鉛含有サブドメインと C 末側サブドメインからなることが分かった。前者は、DNA 結合転写因子である GATA-1 や GRE の zinc-finger と同様の亜鉛配位や局所水素結合網をもつが、亜鉛含有サブドメイン全体は新規で安定な構造をとっていた。後者も、新規な折りたたみを持ち、特徴として正に荷電した大きなくぼみをもっていた。両サブドメインの間には疎水性残基による相互作用が存在した。次に、RPA の単鎖 DNA 結合ドメイン (RPA70 サブユニットの 181-422 残基からなるドメイン)、あるいは、cisplatin による損傷 24-mer 二本鎖 DNA が中央ドメインに結合する部位を XPA の化学シフトの変化により同定した。その結果、亜鉛含有サブドメインは RPA70₁₈₁₋₄₂₂ と相互作用し、C 末側サブドメインのくぼみは DNA と相互作用することが分かった。最後に、アミド ^{15}N 核の緩和時間の測定と解析から、中央ドメインの運動性を解析した。その結果、分子全体の溶液内での回転拡散には異方性が存在し (回転拡散テンソルの主値の比率が 1.38)、その異方性も考慮した内部運動の解析から C 末側サブドメインのくぼみ領域が非常に柔軟であることが分かった。そのくぼみの柔軟性は、XPA 蛋白質が構造の異なるさまざまな DNA 損傷に適応して結合し、それらを認識するために必要であると考察される。

This thesis was approved as to
the style and content by

Professor Yoshimasa Kyogoku

Professor Kiyoji Tanaka

Professor Yuji Goto

Professor Hideo Shinagawa

November 16, 1998.

CONTENTS

ABBREVIATIONS	6
SUMMARY	8
INTRODUCTION	9
EXPERIMENTAL PROCEDURES	13
● Background of NMR Experiments	13
Introduction of stable isotopes into proteins	13
Resonance assignments	13
NMR parameters for structure determinations	14
Structure calculations	14
Protein-ligand interactions	15
Relaxation parameters	15
Model-free analysis	16
● Sample Preparations of XPA ₉₈₋₂₁₉	17
Expression check	17
Non-labelled proteins	17
Uniformly ¹⁵ N-labelled proteins	17
Uniformly ¹⁵ N- and ¹³ C-doubly labelled proteins	18
Specifically ¹⁵ N-Lys or ¹⁵ N-Arg labelled proteins	18
Fractionally deuterated proteins	18
Purification	19
● Sample Preparations of RPA70 and Damaged DNA	20
Preparation of RPA70 ₁₈₁₋₄₂₂	20
Purification of oligonucleotides	20
Platination reaction	21
● NMR Spectroscopy	21
Parameters	21
Resonance assignments	21
Processing of the NMR data	23

Structure determination	24
● Structure Calculation	24
● Chemical Shift Perturbation Experiments	25
● Relaxation Analyses	25
Measurements of ^{15}N T_1 , T_2 , and NOE	25
Estimation of ^{15}N T_1 , T_2 , and NOE	26
$J(0)$ analysis	27
Model-free analysis with an isotropic rotational diffusion model	27
Model-free analysis with an anisotropic rotational diffusion model	28
RESULTS	29
● Preparation of XPA ₉₈₋₂₁₉	29
● Resonance Assignments	31
● Structures	35
Overall structure	35
Zinc-containing subdomain	39
C-terminal subdomain	41
● DNA and RPA70 Binding Surfaces	43
Interaction with DNA	43
Interaction with RPA70	44
● Dynamics	46
^{15}N relaxation data	46
$J(0)$ analysis	48
Isotropic, axial symmetric, fully anisotropic models	49
Model-free analyses	50
Initial estimation of τ_m , and D	50
Selection of the dynamic models	52
Optimization of τ_m , D , and internal motion parameters	53
DISCUSSION	57
● Description of the Structure	57
Comparison with other structures	57
Exons/introns and tertiary structure elements	57

● Interactions of XPA with DNA and RPA70	59
Interaction with damaged DNA	59
Specificity to damaged DNA	59
Interaction with RPA70	60
● Protein Dynamics	62
Selection of the diffusion models	62
Comparison of the dynamic parameters	62
Backbone dynamics and their biological implications	64
CONCLUSION	66
Coordinates and other NMR data	67
REFERENCES	68
List of Publications	77
ACKNOWLEDGEMENTS	78

ABBREVIATIONS

C-terminal	<u>c</u> arboxyl- <u>t</u> erminal
CPMG	<u>C</u> arr- <u>P</u> urcell- <u>M</u> eiboom- <u>G</u> ill
CSA	<u>c</u> hemical <u>s</u> hift <u>a</u> nisotropy
DD	<u>d</u> ipole- <u>d</u> ipole interaction
DEAE	<u>d</u> iethyl <u>a</u> mino <u>e</u> thyl
DSS	2,2- <u>d</u> imethyl-2- <u>s</u> ilapentane-5- <u>s</u> ulfonate sodium salt
DTT	<u>d</u> i <u>t</u> hiou <u>r</u> eitol
EDTA	<u>e</u> thylene <u>d</u> iamine- <u>N,N,N',N'</u> - <u>t</u> etraacetic acid
ERCC1	<u>e</u> xcision <u>r</u> epair <u>c</u> ross <u>c</u> omplementing rodent repair deficiency 1
<i>E. coli</i>	<i><u>E</u>scherichia <u>c</u>oli</i>
HEPES	<u>N</u> -2- <u>h</u> ydroxy <u>e</u> thyl <u>p</u> iperazine- <u>N'</u> -2- <u>e</u> thane <u>s</u> ulfonic acid
HMQC	<u>h</u> eteronuclear <u>m</u> ultiple <u>q</u> uantum <u>c</u> orrelation
HSQC	<u>h</u> eteronuclear <u>s</u> ingle <u>q</u> uantum <u>c</u> orrelation
INEPT	<u>i</u> nsensitive <u>n</u> uclei <u>e</u> nhanced by <u>p</u> olarization <u>t</u> ransfer
IPTG	<u>i</u> sopropyl- β - <u>D</u> - <u>t</u> hiogalactopyranoside
monoS	mono <u>s</u> ulpho
nD-	<u>n</u> <u>d</u> imensional-
NER	<u>n</u> ucleotide <u>e</u> xcision <u>r</u> epair
NMR	<u>n</u> uclear <u>m</u> agnetic <u>r</u> esonance
NOE	<u>n</u> uclear <u>O</u> verhauser <u>e</u> ffect
N-terminal	amide-terminal
OD _n	<u>o</u> ptical <u>d</u> ensity at <u>n</u> nm
PCNA	DNA polymerase accessory factor <u>p</u> roliferating <u>c</u> ell <u>n</u> uclear <u>a</u> ntigen
PMSF	<u>p</u> henyl <u>m</u> ethyl <u>s</u> ulfonyl <u>f</u> luoride
RFC	<u>r</u> eplication <u>f</u> actor <u>C</u>
RMS	root <u>m</u> ean <u>s</u> quare
RPA	replication <u>p</u> rotein <u>A</u> (= the single-strand-binding protein: hSSB)
SD	<u>s</u> tandard <u>d</u> eviation
SDS-PAGE	<u>s</u> odium <u>d</u> odecyl <u>s</u> ulfate - <u>p</u> oly <u>a</u> crylamide gel <u>e</u> lectrophoresis
SP	<u>s</u> ulphopropyl
T_1	spin-lattice longitudinal relaxation time

T_2	spin-spin transverse relaxation time
TFIIH	basal <u>t</u> ranscription <u>f</u> actor <u>I</u> I <u>H</u>
Triton X-100	t-octyl phenoxy poly ethoxy ethanol
WATERGATE	<u>w</u> ater suppression by <u>g</u> radient <u>t</u> ailored <u>e</u> xcitation
XP	<u>x</u> eroderma <u>p</u> igmentosum (Greek: <i>xeros</i> , dry + <i>derma</i> , skin)
XPA	<u>x</u> eroderma <u>p</u> igmentosum group <u>A</u>

S U M M A R Y

XPA is a nucleotide excision repair (NER) protein, which is involved in the damage recognition step of NER. XPA binds to other repair proteins, and acts as a key element in NER complex formation. The central domain of human XPA consisting of 122 amino acids from Met98 to Phe219 is responsible for the preferential binding to damaged DNA and to replication protein A (RPA). In this study, the resonance assignments, three-dimensional structure, interactions with DNA and RPA, and dynamic properties of the central domain were studied by means of NMR. Almost all of the resonances of the backbone and sidechain ^1H , ^{15}N and ^{13}C nuclei were assigned. The calculated structures showed that the domain consists of a zinc-containing subdomain with a compact globular structure and a C-terminal subdomain with a positively charged cleft in a novel α / β structure. The spectra of the complexes with RPA70 and with DNA showed that the zinc-containing subdomain binds to RPA70 and the cleft in the C-terminal subdomain binds to DNA. The results of the analyses of ^{15}N relaxation data showed that the central domain exhibits anisotropic rotational diffusion, and most of flexible regions exist on the DNA binding surface in the cleft in the C-terminal subdomain. The flexibility may be involved in the ability of XPA to recognize various kinds of DNA damages.

INTRODUCTION

Ultraviolet (UV) irradiation and some chemical agents produce alterations in DNA that are potentially mutagenic or lethal to cells. Usually, cells efficiently recognize and eliminate lesions through DNA repair processes. Nucleotide excision repair (NER) is the ubiquitous pathway by which a broad spectrum of structurally unrelated DNAs damaged by UV and chemical carcinogens is removed from the genome. The NER pathway repairs damaged DNA by recognizing lesions, excising the oligomer carrying the damaged bases, and synthesizing a repair patch using the opposite strand as the template (1, 75) (Fig. 1). The importance of NER has been highlighted by studies on the human inherited disease, xeroderma pigmentosum (XP), which is characterized by a >1,000-fold higher frequency of skin cancer in sun-exposed areas, and by neurological complications (2) (Fig. 2). Cells from XP patients have defects in NER, and therefore are hypersensitive to UV irradiation. Complementation analyses have identified seven complementation groups (A-G) and a variant form in XP cells (3).

The gene that complements XP group A cells encodes a zinc-finger protein, XPA, composed of 273 amino acids. XPA has been shown *in vitro* to bind preferentially to various kinds of DNA damages such as (6-4) photoproducts and crosslinks caused by UV and chemical carcinogens. It has, therefore, been suggested that XPA is involved in the damage recognition step of NER (4-8). XPA has also been shown to bind directly to many other repair factors: replication protein A (RPA), ERCC1 (excision repair cross-complementing rodent repair deficiency 1) /XPF heterodimer, and TFIIH (transcription factor II H) which contains XPB and XPD as its subunits (9-16). All of these factors are essential for the early steps of NER (1, 17). *In vitro* experiments have shown that XPA has a moderately higher affinity for damaged DNA over undamaged DNA, and exhibits enhanced preferential binding in the presence of RPA (9, 10). It was also shown that the binding activity of XPA to damaged DNA was increased by the interaction of XPA with ERCC1 (16). Thus, it has been suggested that XPA plays a role in loading the incision protein complex onto a damaged site as a multifunctional protein that coordinates the early steps of NER processes (11, 13). Recently, it has also been reported that XPA, RPA, TFIIH, XPC-HHR23B, XPG and ATP are required for high specificity DNA-protein complex formation (18). There is another suggestion of a two-step mechanism about damage recognition in which damage is detected by XPC-HHR23B, and then verified by XPA (19).

XPA consists of several distinct functional domains. Its amino-terminal part contains a region (residues 4 to 74) for binding to a subunit of RPA, RPA34, and a region (residues 59 to

97) for binding to ERCC1 (11, 12, 14-16). The carboxyl-terminal part of XPA, residues 226 to 273, has been shown to bind to TFIIH and to recruit it to the damaged site (12, 13). The central domain (residues Met98 to Phe219, Fig. 3), designated as MF122 by Kuraoka *et al.*, has been identified as the minimal polypeptide essential for the preferential binding to damaged DNA, by means of a combination of limited proteolysis and deletion analysis (8). The central domain also includes a region essential for binding to RPA70 (10,11).

The central domain contains the zinc-binding sequence, Cys-X-X-Cys-(X)₁₇-Cys-X-X-Cys, in its N-terminal part (20). Through ¹¹³Cd-NMR analysis combined with site-directed mutagenesis and atomic absorption, it has been shown that one zinc ion is tetrahedrally coordinated by four Cys residues in the zinc-binding sequence (20). Miyamoto *et al.* have shown that a mutation at each of the four zinc-coordinating Cys residues results in a drastic reduction in the UV resistance of cells, as compared to the wild-type XPA protein (21). However, a mutation at Cys153, which is located outside of the zinc-binding sequence, gives almost the same UV resistance as that of the wild-type XPA. These results indicate that XPA has a (Cys)₄ type zinc-finger motif, and it was generally thought that the zinc-finger motif of XPA was involved in DNA binding, through analogy to the zinc-finger motifs found in transcriptional regulatory factors.

In this thesis, I present NMR studies of the resonance assignments of the backbone and sidechain ¹H, ¹⁵N, and ¹³C nuclei, solution structure, interactions with damaged DNA and RPA70, and overall and internal dynamics of the central domain of human XPA. The purpose of these studies is to understand the mechanisms of the damage recognition and complex formation with another protein in terms of the three-dimensional structures. The determined structure shows that the central domain consists of a zinc-containing subdomain, a C-terminal subdomain with a basic cleft, and a linker sequence which connects these two subdomains. The spectra of its complex with a cisplatin-damaged oligonucleotide or a truncated RPA70 show that the cleft in the C-terminal subdomain serves as the DNA-binding surface and the zinc-containing subdomain serves as an RPA70-binding surface. The backbone dynamics were determined on analysis of the ¹⁵N *T*₁, *T*₂ and NOE (nuclear Overhauser effect) relaxation data. The results show an anisotropic character of the rotational diffusion of the central domain. The DNA-binding surface exhibits a highly dynamical feature, suggesting possible relationships to the versatility of the damage recognition of XPA.

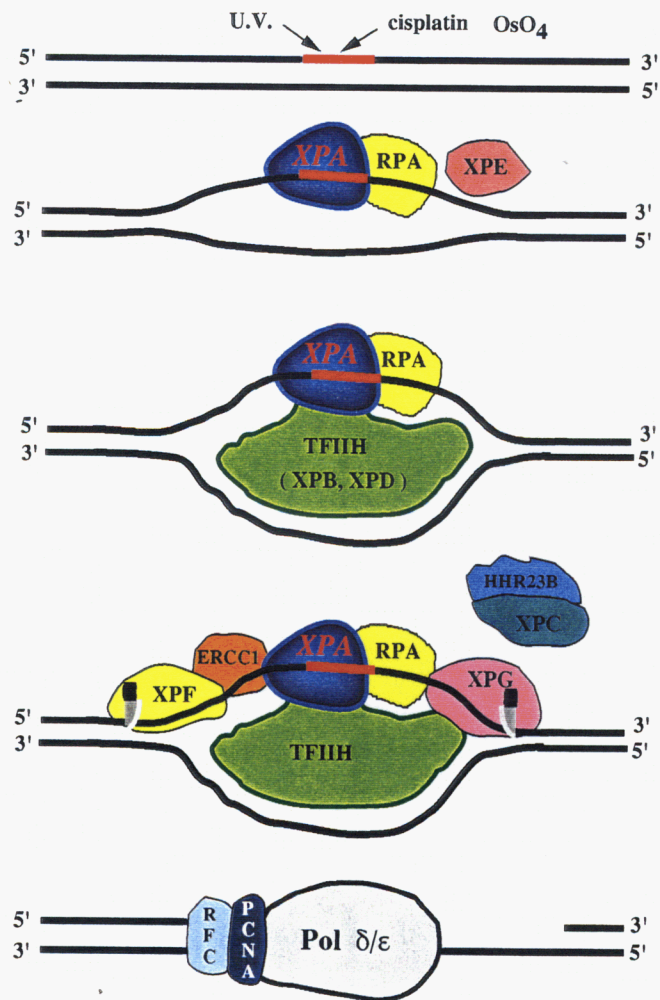
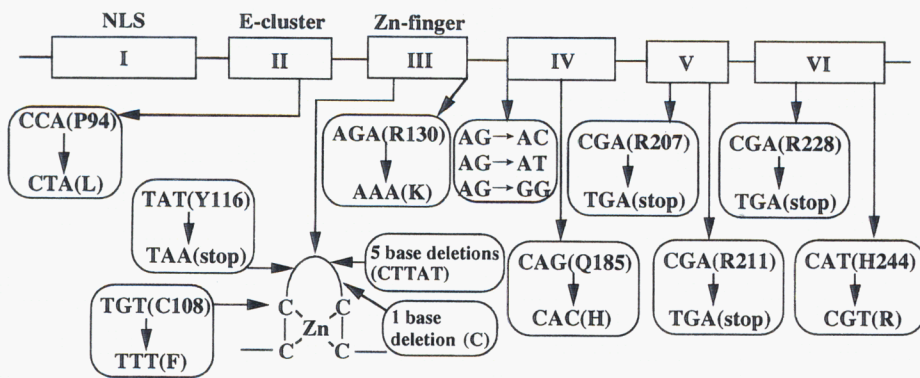


Fig. 1. Model of human nucleotide excision repair (NER).

U.V., cisplatin, OsO_4 , oxygen radical, and so on damage DNA. A damaged site is drawn in red. The XPA-RPA and, possibly, other factors recognize the damaged DNA, and recruit TFIIH to the damaged site. The subunits of TFIIH, XPB and XPD, have ATP-dependent helicase activities with 3' to 5' and 5' to 3' directions, respectively, which may facilitate the formation of an open DNA complex. The role of XPE and XPC-HHR23B are presently unclear. XPG may be recruited to the lesion by its interaction with RPA. Likewise, ERCC1-XPF(ERCC4) complex may be recruited to the lesion by the interaction between ERCC1 and XPA. About 28-mer oligonucleotide including the damage is cut out by XPG and XPF on the 3'- and 5'-sides, respectively. The DNA polymerase δ/ϵ synthesizes a new DNA strand. The proliferating cell nuclear antigen (PCNA) and replication factor C (RFC) may be required to displace the incision proteins from the postincision complex. Finally, DNA ligase I ligates the DNA (75).



Symptoms Acute skin cancers on sun-exposed areas
 High frequency of organ cancers
 Neurological complications

Fig. 2. Mutations found in xeroderma pigmentosum group A patients.

The deletions and base substitutions are indicated in boxes. The main symptoms are also listed. NLS means nuclear localization signal or nuclear transport signal. E-cluster means glutamic acid cluster. The numbered boxes I to VI stand for respective exons.

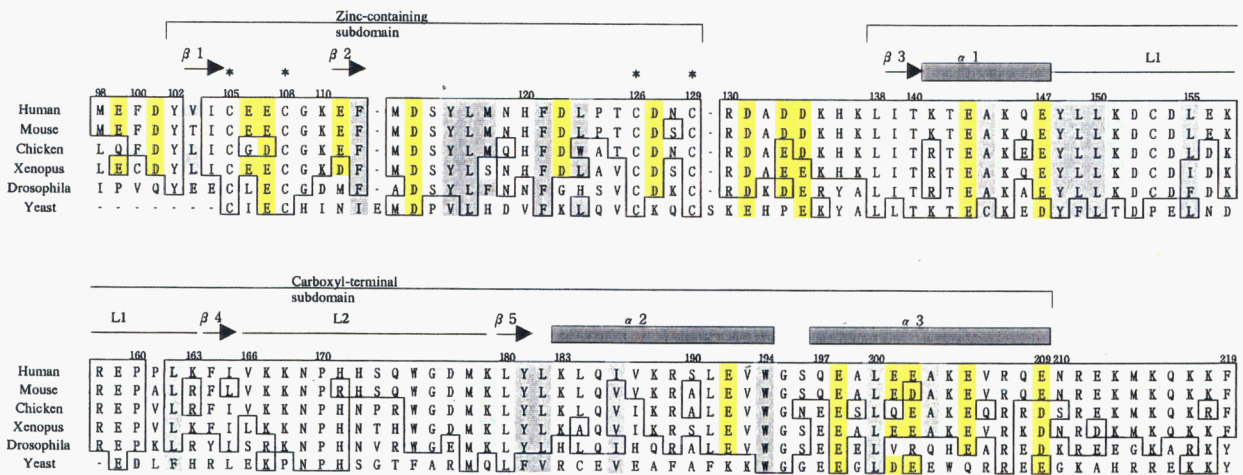


Fig. 3. Sequence alignment of the central domains of human XPA and other XPAs.

The numbering is shown for human XPA. The asterisks indicate the zinc-coordinated Cys residues. The residues identical to human XPA are boxed. The secondary structure of human XPA which was determined in this study is indicated. Important residues discussed in the text are colored (green, basic; yellow, acidic; grey, hydrophobic).

EXPERIMENTAL PROCEDURES

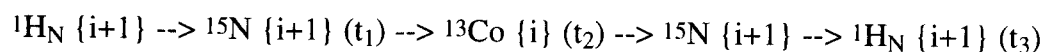
● Background of NMR Experiments

Introduction of stable isotopes into proteins

Assignment of the spin systems in protein NMR spectra is an essential step in solution structure determination. However, conventional proton NMR experiments are ineffective for structure determinations of proteins with molecular masses greater than approximately 10 kDa. The number of protons present in proteins increases with molecular masses. The rotational correlation times of proteins, and thus the linewidths of peaks in NMR spectra, also increase with molecular masses. The increased numbers and linewidths of resonances in homonuclear ^1H NMR spectra result in severe resonance overlaps. The introduction of isotopically (^{15}N - and/or ^{13}C) labelled proteins has enabled heteronuclear NMR spectroscopy which effectively circumvents these problems. Such labelling has become available by the device of high-yield expression systems for proteins in bacteria such as *Escherichia coli*, which can grow in minimal nutrient media composed of simple organic carbon sources and nitrogen sources. Spectral resolution is improved by an increase of the dimensionality of the NMR spectrum. Moreover, the efficiency of coherence transfer is increased by relatively large scalar coupling interactions rather than the small ^1H homonuclear three-bond scalar coupling interactions. Recently, even larger proteins consisting of more than 200 amino acids have become accessible by the introduction of fractional deuteration. That is because the deuteration reduces the ^1H relaxation rates which are otherwise dominated by dipolar interactions with other protons, and the deuteration also reduces the heteronuclear relaxation rates which are otherwise dominated by strong dipolar interactions with directly bonded ^1H nuclei (23, 76, 77).

Resonance assignments

Through-bond connectivities between chemically remote pairs of magnetic nuclei can be established by a stepwise relay of magnetization along a chain of chemically linked magnetic nuclei by utilizing relatively large one-bond scalar coupling constants among hetero atoms. A large number of heteronuclear pulse schemes have been published which provide correlations between backbone and sidechain nuclei. The nomenclature established for these experiments is systematic. For example, HNC0 experiment uses the following coherence transfers:



, where $\{i\}$ represents the amino acid number, i . Magnetizations are detected in the t_1 , t_2 and t_3 time domains, and transformed to the ω_1 , ω_2 and ω_3 frequency domains by Fourier transformations, respectively. Thus, the HNCO experiment provides a cross peak at the cross point of the chemical shifts of the $^1\text{H}_\text{N}/^{15}\text{N}$ spins of residue $i+1$, and the ^{13}C spin of residue i . On the other hand, HN(CA)CO experiment correlates the chemical shifts of the $^1\text{H}_\text{N}/^{15}\text{N}$ spins of residue i , and the ^{13}C spin of the same residue i . Therefore, if the $^{13}\text{C}\{i\}$ shift is unique, the $^1\text{H}_\text{N}/^{15}\text{N}\{i+1\}$ and $^1\text{H}_\text{N}/^{15}\text{N}\{i\}$ shifts are linked. In similar ways, almost all of the chemical shifts of ^1H , ^{13}C and ^{15}N nuclei in proteins can be assigned (23, 76, 77).

NMR parameters for structure determinations

NMR methods for protein structure determinations mainly rely on a dense network of distance constraints derived from nuclear Overhauser effects (NOEs). The NOE reflects the transfer of magnetization between hydrogen atoms separated by less than 5 Å which may be far away along the protein sequence but close together in space. The second source of constraints obtained by NMR is the information on torsion angles derived from scalar coupling constants. Dihedral angle ϕ constraints are often derived from $^3J_{\text{HNH}\alpha}$ coupling constants between a backbone amide proton and the α proton in the same residue according to the Karplus relation ($^3J_{\text{HNH}\alpha}(\phi) = 6.4 \cos^2 \phi - 1.4 \cos \phi + 1.9$). Additional NMR parameters that are often used in structure determinations include hydrogen exchange data and chemical shifts. The presence of slow exchange of amide protons with solvent deuterons is a good indicator of hydrogen bonding patterns. It has been recognized that the deviations of $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^{13}\text{C}\gamma$, and $^1\text{H}\alpha$ chemical shifts from their random coil values are correlated with the local backbone conformations. For example, $^{13}\text{C}\alpha$ chemical shifts larger or smaller than the random coil values tend to occur for α -helix or β -sheet conformation, respectively (23, 76, 77).

Structure calculations

NMR based structure calculations essentially involve finding conformations of a molecule that are consistent with defined structural constraints such as interproton distance, dihedral angle, and hydrogen bonding constraints. There are a variety of methods which utilize restrained molecular dynamics (rMD) and/or distance geometry (DG) methods. The DG uses the metric matrix algorithm or the variable target function approach. The rMD uses the simulated annealing (SA) procedure in which pseudoenergy potentials drive a structure toward a conformation that will

reduce violations of constraints during a forced heatup-cooldown annealing cycle. The result of an NMR structure calculation is a group of conformers that represents the solution structure of a protein. These structures are often evaluated with root mean square deviation (RMSD) values which indicate differences between the structures, and distributions of dihedral angles (e.g., Ramachandran plot) (23, 76, 77).

Protein-ligand interactions

Specific interactions between proteins and ligands play important roles in many biological processes. The simplest and most popular approach to detect the interactions by NMR is the chemical shift perturbation experiment. In the experiment, changes of amide chemical shifts in a protein are monitored by ^{15}N - ^1H -HSQC spectra during titration with a ligand. The locations of resonances that undergo chemical shift perturbations are mapped onto the three-dimensional structure of the protein. This mapping exhibits the putative binding site of the protein for the ligand. Since amide chemical shifts are sensitive to many changes in condition (e.g., pH, ionic strength, temperature, etc.), NMR line widths often increase due to the effective molecular weight of a protein and its ligand (74).

Relaxation parameters

Solution NMR spectroscopy is a powerful tool for characterizing protein dynamics. By perturbing the states of the nuclear spin systems and observing their returns to the equilibrium, NMR relaxation experiments provide a sensitive means by which both global (external) and internal motions in proteins can be probed. Heteronucleus, ^{15}N spin, is well suited for relaxation studies because the relaxation mechanism is governed mainly by the chemical shift anisotropy (CSA) and the dipole-dipole interaction (DD) with the attached ^1H nucleus. As a consequence of molecular motion, the CSA and DD create the fluctuations of local fields, which provide the mechanism for spins to exchange energy with the lattice or with each other, resulting in nuclear spin relaxation. In practice, the longitudinal relaxation time (T_1), transverse relaxation time (T_2), and nuclear Overhauser effect (NOE) are measured and analyzed. (The NOE used in relaxation experiments stands for heteronuclear steady state NOE. To the contrary, the NOE used for distance constraints stands for homonuclear transient NOE.) The theoretical expressions for T_1 , T_2 and NOE are

$$1/T_1 = d^2 [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N)$$

$$1/T_2 = d^2/2 [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + c^2/6 [4J(0) + 3J(\omega_N)]$$

$$\text{NOE} = 1 + (\gamma_{\text{H}}/\gamma_{\text{N}}) d^2 [6J(\omega_{\text{H}} + \omega_{\text{N}}) - J(\omega_{\text{H}} - \omega_{\text{N}})] T_1$$

, where $d^2 = \gamma_{\text{H}}^2 \gamma_{\text{N}}^2 (h/2\pi)^2 / (10 r_{\text{HN}}^6)$ and $c^2 = 2 \gamma_{\text{N}}^2 H_0^2 \Delta^2 / 15$; Δ is the chemical shift anisotropy of amide ^{15}N nucleus; H_0 is the strength of a static magnetic field; ω_{N} is the angular resonance frequency of ^{15}N nucleus; γ_{H} and γ_{N} are gyromagnetic ratios for ^1H and ^{15}N nuclei, respectively; h is Planck's constant; and r_{HN} is the length of the amide ^{15}N - ^1H bond. $J(\omega)$ denotes the spectral density function at angular frequency, ω . It can be thought of as a spectrum analyzer which expresses the probability distribution of motional frequency. If a molecule is known to have isotropic rotational diffusion, the overall rotational correlation time, τ_{m} , can be estimated from a trimmed average value of T_1/T_2 which is assumed to be independent of generalized order parameter, S^2 (23, 76, 77).

Model-free analysis

A useful way of expressing $J(\omega)$ in terms of the minimum number of model independent motional parameters was provided by the formalism of Lipari-Szabo (37, 38). According to this formalism,

$$J(\omega) = S^2 \tau_{\text{m}} / (1 + \omega^2 \tau_{\text{m}}^2) + (1-S^2) \tau_{\text{i}} / (1 + \omega^2 \tau_{\text{i}}^2),$$

$$1/\tau_{\text{i}} = 1/\tau_{\text{m}} + 1/\tau_{\text{e}}$$

for isotropic rotational diffusion, and

$$J(\omega) = S^2 \sum_{k=1,3} A_k \tau_k / (1 + \omega^2 \tau_k^2) + (1-S^2) \tau_{\text{i}} / (1 + \omega^2 \tau_{\text{i}}^2),$$

$$1/\tau_{\text{i}} = 2D_{\parallel} + 4D_{\perp} + 1/\tau_{\text{e}},$$

$$A_1 = (1.5 \cos^2 \alpha - 0.5)^2, A_2 = 3 \sin^2 \alpha \cos^2 \alpha, A_3 = 0.75 \sin^4 \alpha,$$

$$\tau_1 = 1/(6D_{\perp}), \tau_2 = 1/(D_{\parallel} + 5D_{\perp}), \tau_3 = 1/(4D_{\parallel} + 2D_{\perp})$$

for axially symmetric anisotropic rotational diffusion

, where τ_{m} is the overall rotational correlation time for molecular tumbling, τ_{e} is the effective correlation time characterizing the internal motions, α is the angle between the NH bond vector and the unique axis of the diffusion tensor, D , and S^2 is the generalized order parameter which is interpreted as describing the amount of spatial freedom of the NH vector as a result of internal motion. The residue-specific parameters for internal motions, S^2 and τ_{e} , and overall parameter, $\tau_{\text{m}}(D)$, for external motions are adjusted to fit observed relaxation data, T_1 , T_2 and NOE. The τ_{m} value can be compared with the value calculated by hydrodynamic theory according to the equation, $\tau_{\text{m}} = \eta V / kT$, where η is viscosity, k is Boltzmann's constant, T is absolute

temperature, and V is volume of the object. Anisotropy of the rotational diffusion can be estimated from the ratio of the principal values of the diffusion tensor, D . The order parameter, S^2 , contains detailed information on the internal motions (23, 76, 77).

● Sample Preparations of XPA₉₈₋₂₁₉

Expression check

The vector plasmids which carried the genes of the central domain of human XPA protein (XPA₉₈₋₂₁₉) under the control of the T7 ϕ 10 promoters were constructed and kindly provided by Kuraoka *et al.* (8). *E. coli* cells of strain BL21(DE3) were transformed with the plasmids. The transformed cells were plated out onto 2 × YT (16 g/L bacto-tryptone, 10 g/L bacto-yeast extract and 5 g/L NaCl) agar plates containing 50 μ g/ml ampicillin. Several colonies were inoculated into 3 ml of 2 × YT medium containing 50 μ g/ml ampicillin. It was incubated at 37 °C with shaking until the OD₆₂₀ reached 0.4. At this point, 0.5 ml of the culture was taken from the culture, and stored with 50% sterile glycerol at -20 °C. IPTG was added into the remaining culture to a final concentration of 4 mM for the induction of the expression of XPA₉₈₋₂₁₉. Two hours after the induction, the expression was checked by SDS-PAGE (15% acrylamide concentration).

Non-labelled proteins

The stored cells were inoculated into 200 ml 2 × YT medium containing 50 μ g/ml ampicillin, 25 μ g/ml chloramphenicol and 20 μ M zinc-acetate, Zn(CH₃COO)₂, and incubated at 37 °C with shaking for about six hours as a pre-culture. The pre-culture was transplanted into the same 2 L medium. It was incubated with an air-bubbling at 37 °C. When the OD₆₂₀ reached 0.5, IPTG was added to a final concentration of 1.0 mM. Two hours after the induction, the culture was harvested by centrifugation. The pellets were washed with 50 ml of 140 mM NaCl solution at 4 °C. The washed pellets were frozen quickly in liquid N₂ and stored at -20 °C.

Uniformly ¹⁵N-labelled proteins

The stored cells were inoculated into 3 ml 2 × YT medium containing 50 μ g/ml ampicillin and 25 μ g/ml chloramphenicol and incubated at 37 °C with shaking for about four hours. When the OD₆₂₀ reached 1.0, 1.0 ml of the culture was centrifuged at 6,000 rpm, and the pellets were washed with 1.5 ml of sterile 140 mM NaCl solution at 4 °C twice for removal of contamination of

non-labelled nitrogen sources from the 2 × YT medium. The washed pellets were planted into 200 ml M9 minimal medium (7.0 g/L Na₂HPO₄, 3.0 g/L KH₂PO₄, 0.5 g/L NaCl, 20 mg/L thymine, 20 mg/L adenosine, 20 mg/L guanosine, 20 mg/L cytidine, 20 mg/L biotin, 20 mg/L thiamine, 1.0 mM MgSO₄, 3.3 μ M FeCl₃, 50 μ M MnCl₂, and 100 μ M CaCl₂) containing 50 μ g/ml ampicillin, 25 μ g/ml chloramphenicol, 20 μ M Zn(CH₃COO)₂, 0.1% glycerol, 4.0 g/L D-glucose and 0.5 g/L ¹⁵NH₄Cl as a sole nitrogen source as a pre-culture. After growth overnight at 30 °C with shaking, the culture was added into the same 2 L M9 medium. It was incubated at 37 °C with an air-bubbling. When the OD₆₂₀ reached 0.5, IPTG was added to a final concentration of 1.0 mM. Four hours after the induction, the culture was centrifuged and the pellets were washed with 50 ml of 140 mM NaCl solution at 4 °C. The washed pellets were frozen quickly in liquid N₂ and stored at -20 °C.

Uniformly ¹⁵N- and ¹³C-doubly labelled proteins

Since the method was almost the same as that of the singly ¹⁵N-labelled sample described above, only the differences are described here. The medium for obtaining the doubly labelled sample was almost the same M9 medium except that it contained 0.5 g/L ¹⁵NH₄Cl and 1.0 g/L [¹³C₆]D-glucose as sole nitrogen and carbon sources, respectively. It did not contain non-labelled glucose or glycerol. When the OD₆₂₀ reached 0.4, IPTG was added to a final concentration of 2 mM. Four hours after the induction, the cells were harvested by centrifugation.

Specifically ¹⁵N-Lys or ¹⁵N-Arg labelled proteins

The method of the preparation of the ¹⁵N-amino acid specifically labelled sample (64) was similar to that of the singly ¹⁵N-labelled sample described above. The differences were that the medium contained non-labelled (cold) NH₄Cl instead of ¹⁵NH₄Cl, and that 100 mg/L of 95% ¹⁵N₂-labelled L-Lysine-2HCl (or 99% ¹⁵N₄-labelled L-Arginine-HCl) was added together with IPTG (final 1 mM) at the time of the induction (OD₆₂₀ ≐ 0.4).

Fractionally deuterated proteins

A uniformly ¹⁵N-labelled and fractionally (80%) deuterated XPA₉₈₋₂₁₉ (C153S) point mutant was expressed as the following. *E. coli* cells (BL21(DE3)) harboring the vector plasmids which carried the genes of the mutant of the central domain of human XPA (XPA₉₈₋₂₁₉ C153S)

under the control of the T7 ϕ 10 promoters were kindly provided by Kuraoka *et al.* (20). The frozen cells stored in 50% glycerol were streaked onto plates of 2 \times YT 80% D₂O agar medium containing 50 μ g/ml ampicillin. Several colonies were inoculated into 3 ml of 2 \times YT 80% D₂O medium containing 50 μ g/ml ampicillin, and incubated at 37 $^{\circ}$ C with shaking. The expression and purification procedures were similar to those described for the preparation of the ¹⁵N-labelled samples, except that the M9 minimal medium contained 80% D₂O / 20% H₂O. In the similar way, a uniformly ¹⁵N- and ¹³C-labelled and fractionally (50%) deuterated XPA₉₈₋₂₁₉ (C153S) point mutant was expressed. In this case, the medium contained 50% D₂O / 50% H₂O, 0.5 g/L ¹⁵NH₄Cl and 1.0 g/L [¹³C₆]_D-glucose.

Purification

The frozen cells obtained from 2 L of 2 \times YT or ¹⁵N-singly labelled culture were suspended in 30 ml of lysis buffer (50 mM Tris-HCl, 400 mM KCl, 5 mM DTT, 20 μ M Zn(CH₃COO)₂, 1 mM PMSF, and pH 7.2 at 4 $^{\circ}$ C). The cells were lysed by sonication for 30 min on ice. The soluble fraction was separated by centrifugation at 20,000 \times g for 40 min at 4 $^{\circ}$ C. For removal of nucleic acids, the soluble fraction was passed through a DEAE-Sepharose (Pharmacia) anion-exchange opened column (bed volume 20 ml) equilibrated previously with the same lysis buffer. The crude substances were salted out from the pass-through fraction with finally 70% saturated ammonium sulfate. The precipitation was suspended in 400 ml of B buffer (50 mM potassium-phosphate (K₂HPO₄-KH₂PO₄), 50 mM KCl, 5 mM DTT, 20 μ M Zn(CH₃COO)₂, and pH 6.1 at 4 $^{\circ}$ C). The solution was loaded onto a SP-Sepharose (Pharmacia) cation-exchange opened column (bed 20 ml) equilibrated previously with the same B buffer. It was fractionated by a linear KCl gradient ranging from 50 mM to 1000 mM. The fractions eluted at the KCl concentration of around 400 mM were collected. The solution was purified further and concentrated with a Mono-S (Pharmacia) cation-exchange packed column (bed 8 ml) under the same condition as in the case of the SP-Sepharose column. The fractions which contained XPA₉₈₋₂₁₉ were applied onto a Sephacryl S-100 gel filtration column (Hiload™ 26/60, Pharmacia) equilibrated previously with the B buffer whose pH value was increased to 7.3. The sample was concentrated to a volume of about 200 μ L with an Amicon concentrator (Centricon-10 10,000 M.W. cut-off) with 4,500 \times g centrifugation. A desired solution (about 10 mL) was added onto the concentrated sample, and it was concentrated again to a desired volume. Samples for NMR measurements typically comprised 1.2 mM protein in 50 mM deuterated Tris-HCl (pH 7.3 at 30 $^{\circ}$ C), 150 mM KCl, 10 mM

deuterated DTT, and 20 μM $\text{Zn}(\text{CH}_3\text{COO})_2$ in 90% $\text{H}_2\text{O}/10\%$ $^2\text{H}_2\text{O}$.

● Sample Preparations of RPA70 and Damaged DNA

Preparation of RPA70₁₈₁₋₄₂₂

RPA₁₈₁₋₄₂₂, which comprised residues 181 to 422 of RPA70 and had a consecutive stretch of six histidine residues (His-tag) at its N-terminus, was prepared as described (56, 57) and kindly provided by Saijo *et al.* The His-tag sequence was removed with human thrombin (2.5 units/mg RPA protein) in a solution containing 50 mM HEPES, 150 mM NaCl, and 10% glycerol at pH 7.4 at 4 °C. After the incubation for 18 hours, the solution was loaded onto a Mono-S packed column (bed 8 mL) equilibrated with a buffer containing 25 mM HEPES, 20 mM KCl, 3% glycerol, 1 mM DTT, 1 mM EDTA, and 0.01% Triton X-100 at pH 7.4. The column was developed with a linear gradient from 20 to 600 mM KCl. The purified RPA₁₈₁₋₄₂₂ had four extra residues, Gly-Ser-His-Met, at its N-terminus, which was checked by an amino-terminal analysis sequencer.

Purification of oligonucleotides

A chemically synthesized 24-mer deoxyoligonucleotide, 5'- TCT TCT TCT GTG CAC TCT TCT TCT -3' (plat-1), and the complementary oligonucleotide (plat-2) were purchased. Each 4,4'-dimethoxy tritylated (DMTr) oligonucleotide was loaded onto a reversed phase C-18 silica gel column (TOSOH TSK-gel ODS-80Ts) equilibrated with a buffer containing 0.1 M triethylamine acetic acid (TEAA) and 5% acetonitrile at pH 7.0. The column was developed with a linear gradient from 5 to 60% acetonitrile. The elution was concentrated and incubated with 80% acetic acid for 15 min at 20 °C for removal of the DMTr protecting groups. The solution was again purified with the same reversed phase C-18 silica gel column. The concentrated elution was loaded onto an anion-exchange column (Resource-Q, Pharmacia, 1 mL) equilibrated with 200 mM KCl solution, and the triethylamine salts attached on the oligonucleotides were converted to the potassium salts. The solution eluted with 1 M KCl solution was applied onto a gel filtration column (HiTrap desalting, Pharmacia, 5 mL X 3) for the desalting. Each dried oligonucleotide was dissolved in a buffer containing 28 mM $\text{KH}_2\text{PO}_4\text{-K}_2\text{HPO}_4$ (pH7.8), 28 mM KCl, 3 mM DTT, 6 μM $\text{Zn}(\text{CH}_3\text{COO})_2$. Equal amounts of both oligonucleotides were mixed and annealed by decreasing the temperature from 95 to 30 °C overnight.

Platination reaction

A plat-1 solution containing 3 mM plat-1 oligonucleotide, 10 mM Tris-HCl (pH 7.4), and 1 mM EDTA was mixed with twice volume of a platination buffer containing 4.5 mM *cis*-diamminedichloroplatinum II (cisplatin), 4.5 mM NaCl, 0.75 mM Na₂HPO₄, and 0.75 mM NaH₂PO₄. The mixed solution containing 1 mM plat-1 and 3 mM cisplatin was incubated in the dark at 37 °C for 16 hours (58). The platination reaction was stopped by the addition of NaCl to a final concentration of 0.5 M. The solution was applied onto an anion-exchange column and a gel filtration column. The plat-1 oligonucleotide damaged with cisplatin and the undamaged plat-2 oligonucleotide were annealed as described above.

● NMR Spectroscopy

Parameters

All NMR spectra for resonance assignments were acquired at 30 °C with a Bruker AMX500, DMX500 or DRX500 NMR spectrometer equipped with a triple resonance (¹H, ¹⁵N and ¹³C) probe and a self-shielded triple axis gradient coil. The sample tube used was a 5 mm ϕ Shigemi silicone-coated micro tube with a flat bottom and a cylindrical adaptor inserted from upward. The proton base frequency (0 ppm) was set to the absolute resonant frequency of DSS dissolved in 90% H₂O/10% D₂O at 30 °C. The base frequencies of ¹⁵N and ¹³C nuclei were set to the values of 0.10132905 (¹⁵N/¹H) and 0.25144952 (¹³C/¹H) times the proton base frequency, respectively (42).

Resonance assignments

For the assignments of the ¹H, ¹⁵N, and ¹³C resonances, a series of two- and three-dimensional experiments (¹⁵N-¹H-HSQC, ¹³C-¹H-HSQC, ¹⁵N-edited NOESY, ¹⁵N-edited TOCSY, HNCA, HNCO, HN(CA)CO, CBCA(CO)NH, CBCANH, HN(CA)HA, HBHA (CBCACO)NH, HBHA(CBCA)NH, C(CO)NH, H(CCO)NH and HCCH-TOCSY) were performed with the ¹⁵N- or ¹³C-labelled protein dissolved in 90% H₂O/10% ²H₂O (23). The ¹³C-¹H-HSQC and HCCH-TOCSY (69) spectra were also acquired with a sample dissolved in 99.8% ²H₂O.

The 2D-¹⁵N-¹H-HSQC experiments included WATERGATE with '3-9-19' composite pulse sequence (65) and Water-flip-back techniques (23). Other NMR experiments in which amide

protons were directly detected included the sensitivity-enhancement and gradient-echo methods (70) on the indirect ^{15}N dimensions. In these pulse sequences, the first gradient pulse was applied at the end of the first INEPT sequence (23) for a duration of 2 ms at a strength of 25 gauss/cm with a sine curve profile. The second gradient pulse was applied during the half end of the constant ^{15}N evolution period for a duration of 2.5 ms at a strength of 40 gauss/cm with a rectangular profile. The last gradient pulse was applied just before the acquisition period for a duration of 250 μs with a rectangular profile. The strength of the third gradient pulse (40.6 gauss/cm) was precisely adjusted relative to the second pulse, so that they formed a gradient echo between the ^{15}N and ^1H nuclei. The delays after all the gradient pulses were set to 400 μs . The 2D- ^{13}C - ^1H -HSQC experiment was also employed with the sensitivity-enhancement indirect detection on the ^{13}C dimension (67). The delays of the first and the second reverse INEPT (23) sequences were set to 900 $\mu\text{s} \times 2$ and 1.8 ms $\times 2$, respectively. The two pulses for the gradient echo were applied with rectangular profiles for durations of 2.4 ms (40.2 gauss/cm) and 600 μs (40.4 gauss/cm), respectively. The former gradient pulse was applied during an additional delay of 2.8 ms $\times 2$ which was inserted into the ^{13}C indirect detection period.

In NMR experiments in which amide protons were directly detected, the spectral widths were set to 6009.615 and 810.938 Hz for the ^1H and ^{15}N dimensions, respectively. The ^1H carrier was set to the frequency of the water resonance (4.7 ppm), and the ^{15}N carrier was set to 119.4 ppm. In acquisition, 512 (^1H) \times 20 (^{15}N) complex points of 32 to 64 scans were collected.

^1H pulses were applied at field strengths of 27.8 and 6.6 kHz for hard and decoupling pulses, respectively. ^{15}N pulses were applied at field strengths of 6.3 and 1.9 kHz for hard and decoupling pulses during the acquisition period, respectively. ^{13}C α (^{13}Co) hard pulses were applied on the resonance of 54.7 (176.7) ppm at such a strength as to excite none of the ^{13}Co (^{13}C α) regions. These strengths were set to 4.0 and 8.8 kHz for 90 and 180 $^\circ$ pulses, respectively. ^{13}C (aliphatic) pulses were applied on the resonance of 43.7 ppm at strengths of 4.3 and 9.6 kHz for 90 and 180 $^\circ$ pulses, respectively, for the similar reason. Off-resonance ^{13}C pulses were applied with sinc profiles with peak amplitudes of 1.2 and 2.5 kHz for 90 and 180 $^\circ$ pulses, respectively, corresponding to a length of 202 μs . The WALTZ-16 sequence (71) was used for all the decoupling.

For each value of the ^{15}N constant evolution period, the N- and P-type coherences were acquired by inverting the sign of the last gradient pulse and the phase of the ^{15}N 90 $^\circ$ pulse at the end of the evolution period. Each frequency modulated data set was stored in separate memory areas. The phase of the ^{15}N 90 $^\circ$ pulse at the beginning of the evolution period and the receiver

were reversed for each increment of the ^{15}N evolution time (70). The other indirect dimensions were acquired by the TPPI-states method (68).

In the H(CCO)NH, C(CO)NH and HCCH-TOCSY experiments, ^{13}C homonuclear mixing time was set to 20.2 ms and the DIPSI-3 sequence (72) with a field strength of 8.1 kHz was used. The ^{15}N -edited TOCSY experiment was modified to the combination of ^1H -TOCSY and ^{15}N - ^1H -HSQC with the WATERGATE and Water-flip-back techniques described above. The ^1H -homonuclear mixing was employed by the DIPSI-2 sequence (72) with a field strength of 8.6 kHz for a duration of 70.1 ms.

Stereospecific assignments of the methyl groups of the Leu and Val residues were achieved with a 15% fractionally ^{13}C -labelled protein dissolved in 99.8% $^2\text{H}_2\text{O}$ as described (25).

Chemical shifts of the ^1H δ / ϵ spins of aromatic residues were assigned by $(\text{H} \beta) \text{C} \beta (\text{C} \gamma \text{C} \delta) \text{H} \delta$ and $(\text{H} \beta) \text{C} \beta (\text{C} \gamma \text{C} \delta \text{C} \epsilon) \text{H} \epsilon$ experiments in which the ^1H δ / ϵ and ^{13}C β resonances were correlated via scalar couplings (66).

Chemical shifts of the amide groups of Lys or Arg residues were assigned with a ^{15}N - ^1H -HSQC spectrum of ^{15}N -Lys or ^{15}N -Arg specifically labelled XPA₉₈₋₂₁₉, respectively.

I developed a four-dimensional HN(CA)NH experiment which provides the correlations between the amide (^1H - ^{15}N) resonances of neighboring residues. A 4D-HN(CA)NH spectrum was obtained with a uniformly ^{15}N - and ^{13}C -labelled and fractionally (50%) deuterated XPA₉₈₋₂₁₉ (C153S) point mutant as described (73). The deuterium decoupling was applied as continuous waves with a field strength of 1.15 kHz at the carrier frequency of 4.6 ppm. The assignments of the backbone amide groups were sequentially confirmed with the spectrum.

Processing of the NMR data

All data were processed with the program, NMRPIPE (31). The peaks were analyzed with the program, PIPP (32). All the programs were run on Unix workstations (Silicon Graphics Indy or Indigo2).

For the measurements which utilized the sensitivity-enhancement indirect detection method, the amplitude modulated data sets were generated by adding and subtracting the frequency modulated N- and P-type data sets and applying a 90° zero-order phase correction for the direct observation axis to one of the new data sets. The data of the dimension in which the constant evolution period was used were elongated by the Mirror-Image-Linear-Prediction method (23).

The data were multiplied with a cosine-squared window function. A sequence of zero values was added at the end of the data. Fourier transformation was applied. The phase of the data was

adjusted so that the pure absorption peaks were obtained. The imaginary part of the data was discarded. A desired region of the spectrum was extracted. The dimensions were exchanged, and the data of another dimension were processed in the same way as described above.

Structure determination

Interproton distance information was obtained from a 2D, 3D- ^{15}N - or ^{13}C -edited or 4D- ^{13}C , ^{15}N - or ^{13}C , ^{13}C -edited NOESY spectrum acquired at 30 °C with a Bruker DRX500 or DRX800 NMR spectrometer with a mixing time of 100 ms (23). In addition, 14 restraints were obtained through a ^{15}N , ^{15}N -edited 4D NOESY experiment on the uniformly ^{15}N -labelled and fractionally deuterated XPA₉₈₋₂₁₉ (C153S) mutant, since the mutant was shown to have an identical conformation with the wild type XPA₉₈₋₂₁₉ based on the ^{15}N - ^1H -HSQC and ^{113}Cd -NMR spectra (20). In the ^{15}N , ^{15}N -edited 4D NOESY experiment, the mixing time and the pulse delay were set to 200 ms and 2 s, respectively, and the WATERGATE and Water-flip-back techniques were used.

Average values of the backbone vicinal coupling constants ($^3J_{\text{HN,H}\alpha}$) estimated from HMQC-J and HNHA experiments were used for torsion angle constraints (23). The torsion angles, χ_1 , of Tyr116 and His136 were estimated from the $^3J_{\text{C}'\text{C}\gamma}$ and $^3J_{\text{NC}\gamma}$ coupling constants (26).

Amide proton exchange with solvent was monitored by means of a series of ^{15}N - ^1H -HSQC spectra after dissolving lyophilized ^{15}N -labeled XPA₉₈₋₂₁₉ in $^2\text{H}_2\text{O}$ to a final concentration of 0.58 mM. The spectra were recorded at pH 6.5 at 30 °C every 30 min. The signals remaining after exchange were identified.

● Structure Calculation

The structures of the central domain of XPA were calculated by the standard simulated annealing protocol (modified sa.inp file) in the program, X-PLOR 3.1 (27). The NOE connectivities derived from strong, medium, and weak cross peaks were categorized, and assumed to correspond to the upper limits for proton-proton distances of 3.0, 4.0, and 5.0 Å, respectively. The constraints for the hydrogen bonds were added for slowly exchanging amides as 2.7 - 3.3 Å for N-O, and 1.7 - 2.3 Å for H-O. Those in α -helices were estimated based on the short and medium range NOE patterns. Those in β -sheets were introduced during the refinement stages.

The geometry of the zinc-binding site was restrained by a series of sulphur-sulphur distance restraints between Cys105, Cys108, Cys126, and Cys129, based on the previous metal binding results (20), so as to ensure the tetrahedral coordination geometry of the Cys-Zn bonds.

The final structures were analyzed with MOLMOL (28), and AQUA and PROCHECK-NMR (29) software. The secondary structures were determined mainly based on the main chain hydrogen bond patterns and the dihedral angles, ϕ and ψ , of the calculated structures, using PROCHECK-NMR and MOLMOL software.

● Chemical Shift Perturbation Experiments

An equimolar amount of the 24-mer oligonucleotide duplex treated with cisplatin was added to 50 μM ^{15}N -labelled XPA₉₈₋₂₁₉ in 77 mM $\text{KH}_2\text{PO}_4\text{-K}_2\text{HPO}_4$ (pH7.8), 77 mM KCl, 7 mM DTT, 15 μM $\text{Zn}(\text{CH}_3\text{COO})_2$, and 5 mM MgCl_2 . The WATERGATE and Water-flip-back ^{15}N - ^1H -HSQC spectrum at 30 $^\circ\text{C}$ was compared with that of the free XPA₉₈₋₂₁₉. A HSQC spectrum of XPA₉₈₋₂₁₉ mixed with an equimolar amount of the duplex not treated with the chemical was also measured.

For analysis of the XPA-RPA70 interaction, an equimolar amount of RPA70₁₈₁₋₄₂₂ was added to 50 μM ^{15}N -labelled XPA₉₈₋₂₁₉ in 28 mM $\text{KH}_2\text{PO}_4\text{-K}_2\text{HPO}_4$ (pH7.8), 28 mM KCl, 3 mM DTT, 6 μM $\text{Zn}(\text{CH}_3\text{COO})_2$, and 5 mM MgCl_2 . The ^{15}N - ^1H -HSQC spectrum was then compared with that of the free XPA₉₈₋₂₁₉.

● Relaxation Analyses

Measurements of ^{15}N T_1 , T_2 , and NOE

Spectra for ^{15}N spin-lattice relaxation times, T_1 , ^{15}N spin-spin relaxation times, T_2 , and $\{^1\text{H}\}$ - ^{15}N steady-state heteronuclear NOE values were acquired at 30 $^\circ\text{C}$ with Bruker DRX500 and DRX800 spectrometers, equipped with pulsed field gradient probes, operated at ^{15}N frequencies of 50.7 and 81.1 MHz, respectively. The sensitivity-enhanced pulse sequences, used for these experiments, were described previously (30). The T_1 relaxation delays were 5, 65, 145, 245, 365, 525, 755 and 1145 ms at 50.7 MHz, and 5, 105, 215, 355, 525, 735, 1035, 1555 and 2000 ms at 81.1 MHz. The T_2 relaxation delays were 14.4, 28.8, 43.2, 57.6, 72.0, 100.8 and

144.0 ms at 50.7 MHz, and 7.2, 14.4, 28.8, 43.2, 57.6, 79.2, 108.0 and 151.2 ms at 81.1 MHz. The delay between ^{15}N 180° pulses in the Carr-Purcell-Meiboom-Gill (CPMG) sequence for the T_2 measurements was 900 μs . In the experiments for $\{^1\text{H}\}$ - ^{15}N steady-state NOE, relaxation delays of 3.6 and 1.6 s before the ^1H saturations of 3.0 and 5.0 s were applied at 50.7 and 81.1 MHz, respectively. The ^1H saturations were achieved with 120° ^1H pulses applied every 5 ms. The spectral widths were 6009.615 Hz (^1H) and 810.938 Hz (^{15}N) at 50.7 MHz, and 13020.833 Hz (^1H) and 1297.387 Hz (^{15}N) at 81.1 MHz. The ^1H carrier was set to the frequency of the water resonance (4.7 ppm), and the ^{15}N carrier was set to 119.4 ppm. For the T_1 and T_2 measurements, 512 (^1H) X 200 (^{15}N) and 1024 (^1H) X 200 (^{15}N) complex points of 32 scans were collected at 50.7 and 81.1 MHz, respectively. For the $\{^1\text{H}\}$ - ^{15}N NOE measurements, 512 (^1H) X 200 (^{15}N) and 1024 (^1H) X 128 (^{15}N) complex points of 64 scans were collected at 50.7 and 81.1 MHz, respectively. All data sets were processed with the program, NMRPIPE (31). Lorentz-to-Gauss transformations were applied to both ^1H and ^{15}N dimensions.

Estimation of ^{15}N T_1 , T_2 , and NOE

Peak heights were determined from the spectra using the program package, PIPP (32). The root mean square (r.m.s.) value of the background noise of each spectrum was used for the uncertainties of the measured intensities. Each T_1 and T_2 value was determined by fitting a series of measured intensities to a two-parameter function of the form, $I(t) = I_0 \exp(-t / T_{1,2})$, where $I(t)$ is the intensity after a time delay t , and I_0 is the intensity at time zero. Nonlinear least-square fitting according to the Levenberg-Marquardt method was employed for optimization of the values of the I_0 and $T_{1,2}$ parameters using the program, CURVEFIT (33). Only the results with χ^2 values smaller than the tabulated χ^2 values at the 95% confidence level were used, where χ^2 equals $\sum \{I_c(t) - I_e(t)\}^2 / \sigma_1^2$ ($I_c(t)$: the intensity calculated from the fitting parameter values, $I_e(t)$: the experimental intensity, σ_1 : the uncertainty of the experimental value), and summation was performed for the available experimental data set. The uncertainties of the $T_{1,2}$ values were estimated from the Levenberg-Marquardt error matrices (34). The $\{^1\text{H}\}$ - ^{15}N steady-state NOE values were determined from the ratios of the intensities of the peaks with and without proton saturation. The uncertainties of the NOE values were estimated according to the error propagation equation (34).

J(0) analysis

The effective spectral density function at zero-frequency, $J_{\text{eff}}(0)$, is defined as the sum of inherent $J(0)$ and a chemical exchange term. The $J_{\text{eff}}(0)$ values for the backbone ^{15}N nuclei were calculated from the ^{15}N T_1 , T_2 , and NOE values according to the following equations (35, 36),

$$J_{\text{eff}}(0) = J(0) + \lambda R_{\text{ex}} = \lambda \left[-1/(2T_1) + 1/T_2 - 3\gamma_{\text{N}}(\text{NOE}-1)/(5\gamma_{\text{H}}T_1) \right],$$

where $\lambda = 3/(6d+2c)$, $d = \gamma_{\text{H}}^2 \gamma_{\text{N}}^2 (h/2\pi)^2 / (4r_{\text{HN}}^6)$ and $c = \Delta^2 \omega_{\text{N}}^2 / 3$; Δ is the chemical shift anisotropy of the amide ^{15}N nucleus (-160 ppm); ω_{N} is the angular resonance frequency of the ^{15}N nucleus; γ_{H} and γ_{N} are gyromagnetic ratios for the ^1H and ^{15}N nuclei, respectively; h is Planck's constant; and r_{HN} is the length of the amide ^{15}N - ^1H bond (1.02 Å). The λ values calculated at 50.7 MHz ($\lambda_{50.7}$) and 81.1 MHz ($\lambda_{81.1}$) are approximately 0.787×10^{-9} and 0.613×10^{-9} (s²/rad), respectively. The chemical exchange rate, R_{ex} , was calculated from the $J_{\text{eff}}(0)$ values obtained at 50.7 MHz ($J_{\text{eff}}(0)_{50.7}$) and 81.1 MHz ($J_{\text{eff}}(0)_{81.1}$) according to the following equation (35, 36),

$$R_{\text{ex}} (\text{at } \omega_{\text{N}}) = \omega_{\text{N}}^2 (J_{\text{eff}}(0)_{81.1} - J_{\text{eff}}(0)_{50.7}) / (\lambda_{81.1} \omega_{\text{N}81.1}^2 - \lambda_{50.7} \omega_{\text{N}50.7}^2).$$

The uncertainties of the $J_{\text{eff}}(0)$ and R_{ex} values were estimated from the uncertainties of the ^{15}N T_1 , T_2 , and NOE values according to the error propagation equation (34).

Model-free analysis with an isotropic rotational diffusion model

An initial estimate of the overall rotational correlation time, τ_{m} , was obtained from the average T_1/T_2 ratios at 50.7 MHz which were limited within one σ (SD). In the model-free analysis developed by Lipari and Szabo (37, 38) with the assumed isotropic rotational diffusion model, the following five dynamic models were applied according to Mandel *et al.* (33). The models and the combinations of optimized parameters were (1) S^2 , (2) S^2 and τ_{ef} , (3) S^2 and R_{ex} , (4) S^2 , τ_{ef} and R_{ex} , and (5) S_{f}^2 , S^2 and τ_{es} , where S^2 is the square of the order parameter, τ_{e} is the effective correlation time, R_{ex} is the chemical exchange rate, and subscripts f and s indicate the fast and slow time scales, respectively. In model 5, the extended formula of the spectral density function was used (39). For the residues of which the ^{15}N relaxation data were available at both 50.7 MHz and 81.1 MHz, model-free analyses were performed by fitting the data simultaneously for the two fields with the assumed quadratic field dependence of the exchange contribution to ^{15}N transverse relaxation rate ($1/T_2$). Otherwise, the analyses were performed by

fitting only the available data. First, the data were fitted to each of the five dynamic models with the overall rotational correlation time, τ_m , fixed to the initially estimated value. Then, one of the dynamic models was statistically selected for each backbone ^{15}N spin by means of Monte Carlo numerical simulations and F-statistical tests as described by Mandel *et al.* (33). Finally, τ_m and the internal motion parameters were simultaneously optimized with the selected dynamical models. The uncertainty for each parameter was estimated by means of a Monte Carlo simulation. All the analyses were performed using the program, MODELFREE-4.0 developed by Palmer *et al.* (33).

Model-free analysis with an anisotropic rotational diffusion model

The magnitude and orientation of the anisotropic rotational diffusion tensor, D , were estimated from the T_1/T_2 ratios and the molecular coordinates of the central domain of XPA. Amide groups were excluded which had significant internal motions manifested by $\{^1\text{H}\}$ - ^{15}N steady-state NOE values of less than 0.65 or had large exchange rates manifested by R_{ex} values estimated on $J(0)$ analysis of larger than 2.5 s^{-1} at 50.7 MHz . The structure which showed the lowest backbone r.m.s.d. from the mean of the final 30 structures calculated through the simulated annealing procedure was used for the analysis (PDB: 1xpa). The axially symmetric diffusion tensor was calculated with the program, R2_R1_DIFFUSION, in the package, MODELFREE-4.0 (33), and the fully anisotropic diffusion tensor was calculated by using an in-house program. In the analyses, the experimentally obtained T_1/T_2 ratios were fitted to an anisotropic diffusion model by means of nonlinear optimization procedures, so that they agreed best with the T_1/T_2 ratios predicted from the fittings (40). Based on the estimated magnitude and orientation of the anisotropic rotational diffusion tensor, D , model-free analysis with the axially symmetric anisotropic diffusion model was performed. The procedure was essentially the same as that employed in the analysis with the isotropic diffusion model except for the following points. First, the diffusion tensor, D , was used instead of the single overall rotational correlation time, τ_m . Second, the extended expression of the spectral density function (model 5) was not used in the model selection procedure. Third, the information regarding the orientations of the amide NH bond vectors was used, which was obtained from the molecular coordinates. Last, the spectral density function for an axially symmetric diffusion model (40, 41) was applied.

RESULTS

● Preparation of XPA₉₈₋₂₁₉

The central domain of human XPA protein consists of 122 amino acid residues ranging from Met98 to Phe219. The molecular mass of the domain is 14,685 dalton. The domain contains 5 Arg, 18 Lys, 4 His, 10 Asp, 15 Glu, 4 Tyr, 5 Cys, and 2 Trp. The calculated isoelectric point is 6.1. The molar absorption coefficient at 280 nm is calculated to be 17,430 according to the equation $((5,500 \times n [\text{Trp}] + 1,400 \times n [\text{Tyr}]) \times 1.05)$, where $n [x]$ represents the number of amino acid x . A protein concentration was estimated with the coefficient.

A two-dimensional (2D) ^{15}N - ^1H -HSQC experiment provides a spectrum which correlates the resonances of an amide proton ($^1\text{H}_\text{N}$) and the directly attached amide nitrogen (^{15}N). Each cross peak corresponds to the amide group of each amino acid or that of the side chain of Asn, Gln, Trp, His, Lys, or Arg residue. Therefore, a condition of a protein can be judged from the degree of the dispersions and intensities of peaks in HSQC spectra. Comparisons of the HSQC spectra obtained at various conditions of XPA₉₈₋₂₁₉ solutions showed that the domain is more stable when the pH value is higher (Fig. 4), the concentration of salt is higher, the concentration of the domain is lower, or the temperature is lower. However, a lower pH value, lower salt concentration, higher protein concentration, and higher temperature are generally preferable for NMR measurements. Consequently, the final condition was decided to be 1.2 mM protein, 50 mM Tris-HCl (pH 7.3 at 30 °C), and 150 mM KCl, which satisfies both of the protein stability and NMR measurements to some extent.

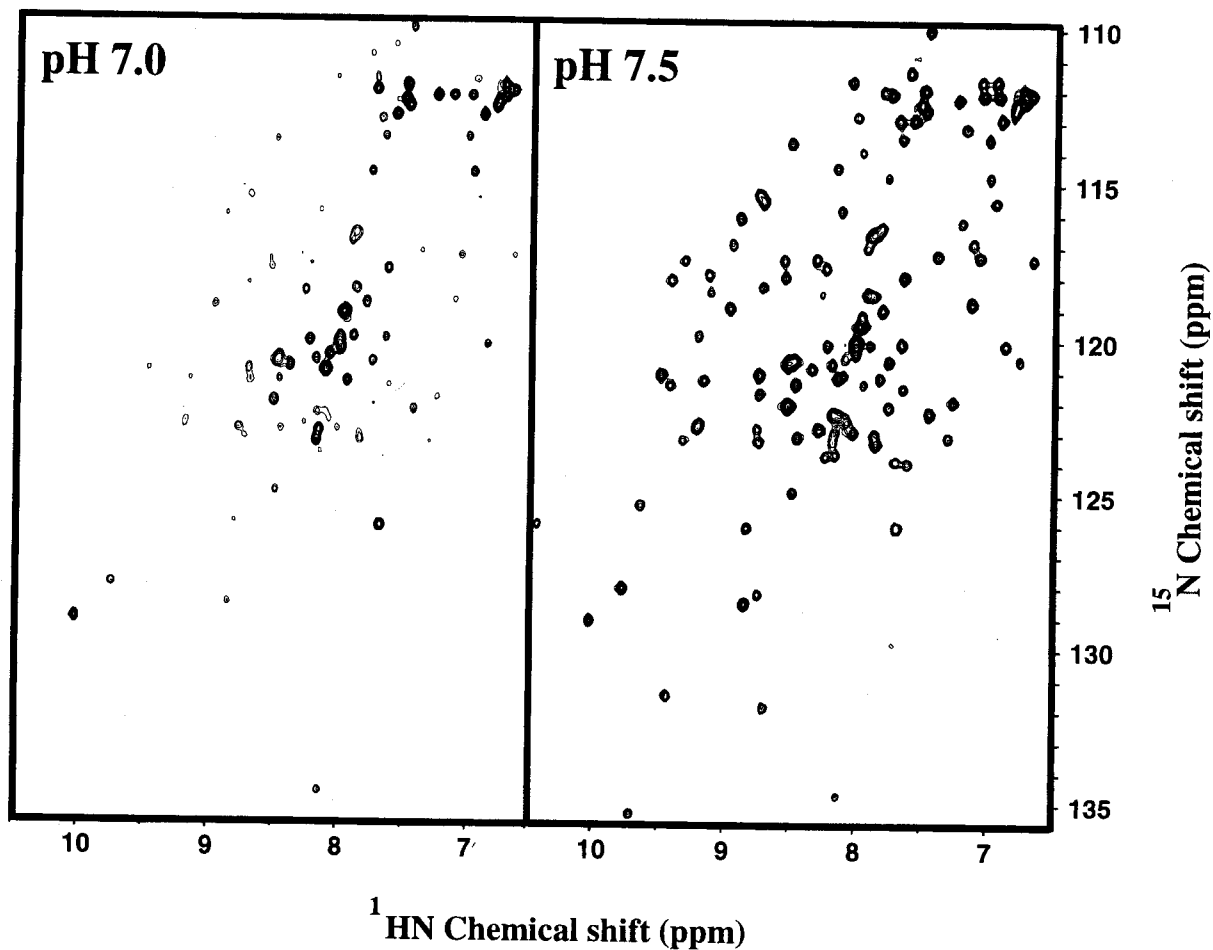


Fig. 4. A pH dependence of the stability of XPA₉₈₋₂₁₉.

Two-D-¹⁵N-¹H HSQC spectra of XPA₉₈₋₂₁₉ obtained at pH 7.0 (left) and 7.5 (right) at 25°C for 1.5 hours by the AMX500 NMR spectrometer. The solution contained 20 mM KH₂PO₄-K₂HPO₄, 180 mM KCl, 20 μM Zn(CH₃COO)₂, and 0.61 mM protein. The inconsistent intensities of the cross peaks in the left spectrum indicate that the protein is not stable at pH 7.0. On the other hand, almost the same intensities of the cross peaks in the right spectrum indicate that the protein is more stable at pH 7.5 than at pH 7.0.

● Resonance Assignments

The resonance assignments for the backbone nuclei were completed except for those of residues 98, 99, 131, 152, 153, and 170-174. The $^{15}\text{N}/^1\text{H}$ signals of the backbone amides of these residues, except for Pro170, could not be observed, probably due to conformational averaging or hydrogen exchange with solvent. Most of the side-chain resonances were assigned except for those of residues 151, 152, and 170-173. All the methyl groups of the twelve Leu and five Val residues were assigned stereospecifically with a 15% fractionally ^{13}C -labelled protein. Fig. 5 shows the sequential connectivities of backbone amide resonances based on $^{13}\text{C}_\alpha$ resonances in HNCA spectrum as an example. Fig. 6 shows a ^{15}N - ^1H -HSQC spectrum of XPA₉₈₋₂₁₉ with the assignments of the backbone and sidechain amide groups.

A four-dimensional HN(CA)NH spectrum was obtained with a uniformly ^{15}N - and ^{13}C -labelled and fractionally deuterated XPA₉₈₋₂₁₉ (C153S) point mutant (73). The assignments of the backbone amide groups were easily confirmed with the spectrum. The experiment provides the correlations between the amide proton and nitrogen resonances of neighboring residues. Thus, the spectrum is extremely beneficial for determining sequential connectivities between neighboring mainchain amide groups. In the experiment, magnetization on $^{15}\text{N}(i)$ spin is at first relayed to $^{13}\text{C}_\alpha(i)$ and $^{13}\text{C}_\alpha(i-1)$ spins by $1J[^{15}\text{N}(i)-^{13}\text{C}_\alpha(i)]$ and $2J[^{15}\text{N}(i)-^{13}\text{C}_\alpha(i-1)]$ scalar interactions, respectively. During the following constant time period, antiphase $^{13}\text{C}_\alpha$ magnetization refocuses partly with respect to $^{15}\text{N}(i)$, while the couplings to the ^{15}N spins of the following and preceding residues evolve. Finally, the magnetization is transferred to the $^{15}\text{N}(i-1)$ and $^{15}\text{N}(i+1)$ spins, as well as back to the $^{15}\text{N}(i)$ spin. Therefore, two cross peaks and one diagonal peak can be observed which correspond to connectivities to $^1\text{H}-^{15}\text{N}(i-1)$, $^1\text{H}-^{15}\text{N}(i+1)$, and $^1\text{H}-^{15}\text{N}(i)$, respectively. For the increase in signal intensities, hydrogens at α sites of a sample are replaced by deuteriums because the transverse relaxation of $^{13}\text{C}_\alpha$ magnetization due to dipolar interactions is reduced by the deuteration. Further, deuterium decoupling is used in the constant time period for the removal of the scalar relaxation of the second kind for $^{13}\text{C}_\alpha$ spins caused by the attached deuterons.

Fig. 7 shows the sequential NOE connectivities, amide proton exchanges, $^3J_{\text{HN},\text{H}\alpha}$ coupling constants, and chemical shift indexes (CSI) for $^{13}\text{C}_\alpha$, $^1\text{H}_\alpha$ and $^{13}\text{C}'$ resonances (42).

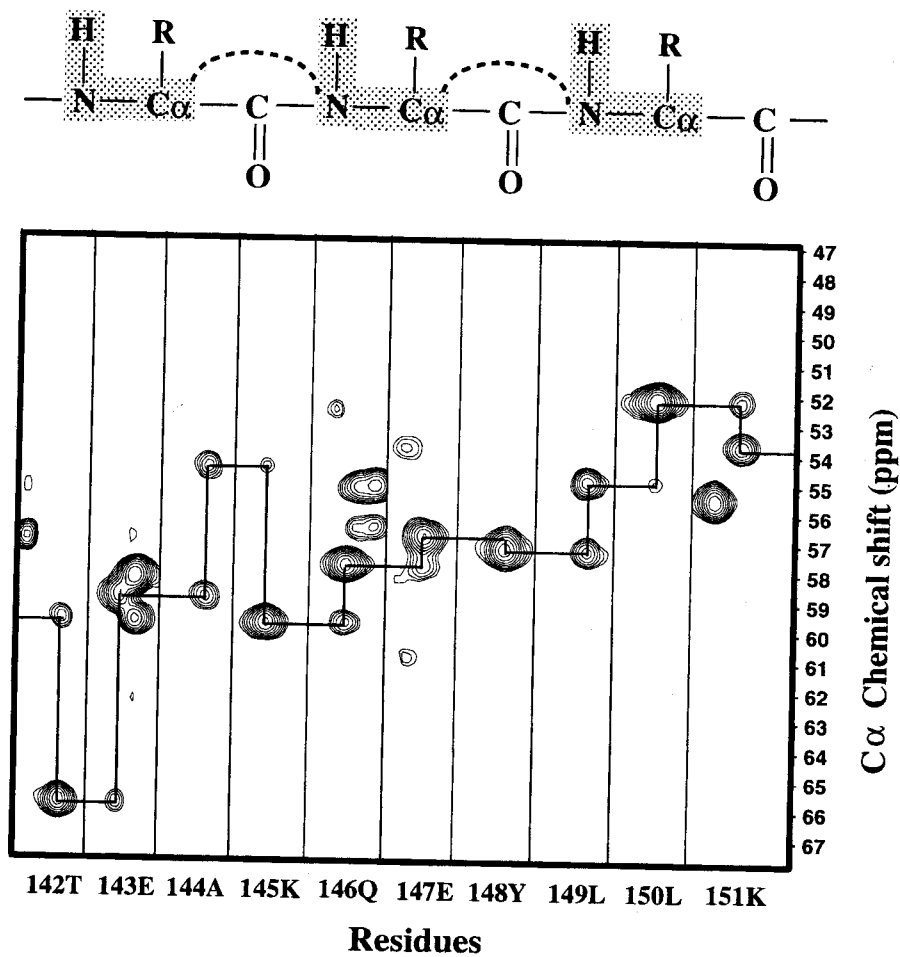


Fig. 5. Strips from HNCA spectrum of XPA₉₈₋₂₁₉.

The strips are taken from slices at the backbone amide ^{15}N frequency of each residue. The HNCA experiment correlates amide ^1H and ^{15}N chemical shifts with $^{13}\text{C}_\alpha$ shifts of the intraresidue as well as the preceding residue. Sequential connectivities of the amide groups are indicated by horizontal lines. Connectivities between cross peaks to $^{13}\text{C}_\alpha$ resonances of residue i and residue $i-1$ are indicated by vertical lines. Interresidual cross peaks generally have weaker intensities than intraresidual cross peaks due to smaller ^{15}N - $^{13}\text{C}_\alpha$ 2J coupling constants than 1J coupling constants.

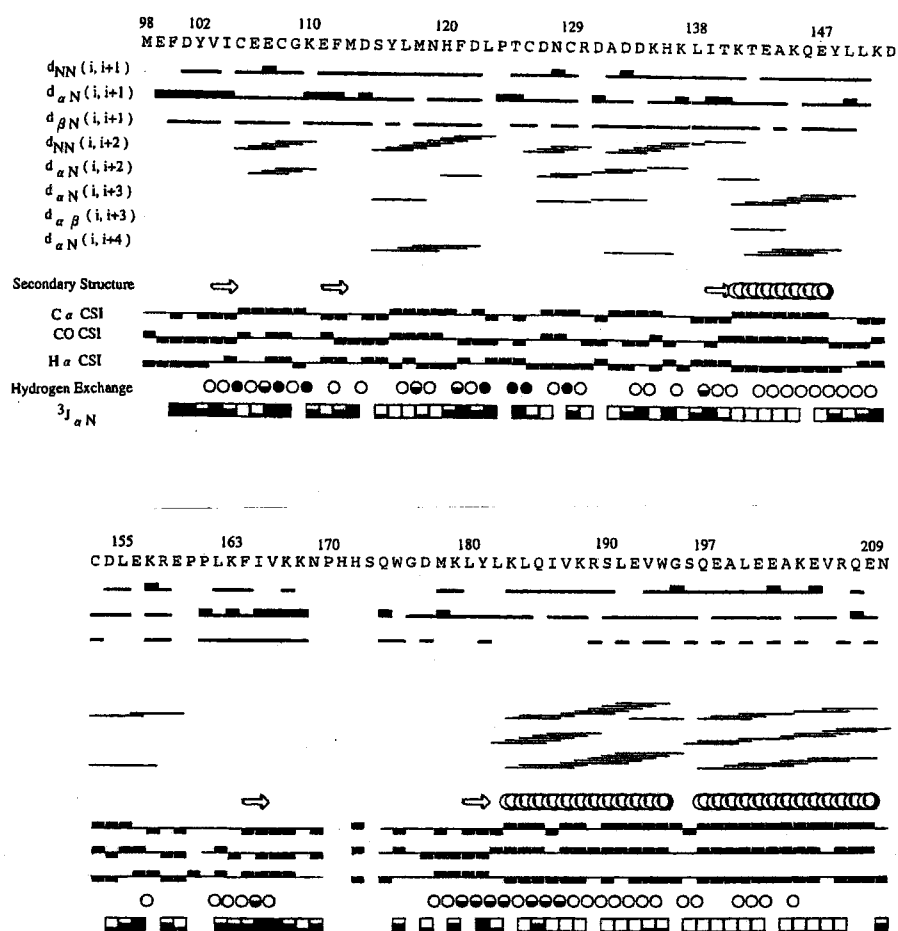


Fig. 7. Summary of the sequential and medium-range NOE connectivities, secondary structures, chemical shift indexes, amide hydrogen exchange data, and $^3J_{\text{HN,H}\alpha}$ coupling constants observed for the central domain of human XPA.

The NOE connectivities are represented by bars, the size of which indicates the NOE intensity (strong or weak). The notation $d_{\alpha\text{N}}(i, i+1)$, for example, shows the connectivity between the α proton resonance of an amino acid (i) and the amide proton resonance of the subsequent amino acid ($i+1$) in the sequence. Amide protons exchanged slowly at pH 6.5 at 30°C are indicated. (Open circles) Life time longer than 30 min and shorter than 2 h. (Half filled circles) Life time longer than 2 h and shorter than 10 h. (Filled circles) Life time longer than 10 h. Three-bond scalar coupling constants between spins H_N and H_α ($^3J_{\alpha\text{N}}$) smaller than 4 Hz are indicated by open boxes, those larger than 4 Hz and smaller than 6 Hz are indicated by one-third filled boxes, those larger than 6 Hz and smaller than 8 Hz are indicated by two-thirds filled boxes, and those larger than 8 Hz are indicated by filled boxes. The chemical shift indexes (CSI) are plotted for $^{13}\text{C}_\alpha$, $^1\text{H}_\alpha$ and $^{13}\text{C}'$ resonances (42). Upper bars represent + 1, lower bars -1, and horizontal lines 0. The figure was produced with the program, VINCE (1996, Rowland Institute for Science).

● Structures

Overall structure

The structure calculations were based on the simulated annealing procedure with 1,336 NOE-derived interproton distance constraints (Fig. 8) (293 intraresidual, 376 sequential, 286 medium-range, and 381 long-range), 83 dihedral angle constraints, 47 sets of hydrogen bond constraints, and six sulphur-sulphur constraints. A total of 140 structures was calculated, and 47 of them showed no violation greater than 0.5 Å or 5°. Of these, the 30 final structures which showed the lowest energy and no violation greater than 0.3 Å or 5° were selected and analyzed (Figs. 9, 10a). The r.m.s.d. value from the average coordinate positions was 0.76 Å for the backbone heavy atoms of residues 102 to 155, 163 to 165, and 180 to 209 (22). A summary of the constraints and structural statistics is listed in Table I. The linearity of the 41 hydrogen bonds of the 30 final structures, to which backbone hydrogen bond constraints were applied, was checked. For the total 1,230 pairs the average and SD values of the O-H_N distance and the angles between N-H_N and N-O, and between C'=O and C'-H_N, are 2.21 Å ± 0.11, 23.9° ± 9.8, and, 27.5° ± 13.7, respectively.

The central domain of XPA consists of a zinc-containing subdomain (residues 102 to 129) and a C-terminal subdomain (residues 138 to 209), connected by an eight amino acid linker sequence (Fig. 10b). Since 42 NOE contacts were found between the zinc-containing subdomain, the C-terminal subdomain, and the linker sequence, the relative orientation of the two subdomains, as well as that of the linker sequence, are well defined. Since few medium- and no long-range NOEs were observed for the C-terminal sequence (residues 211 to 219), this region was excluded from the structure calculations.

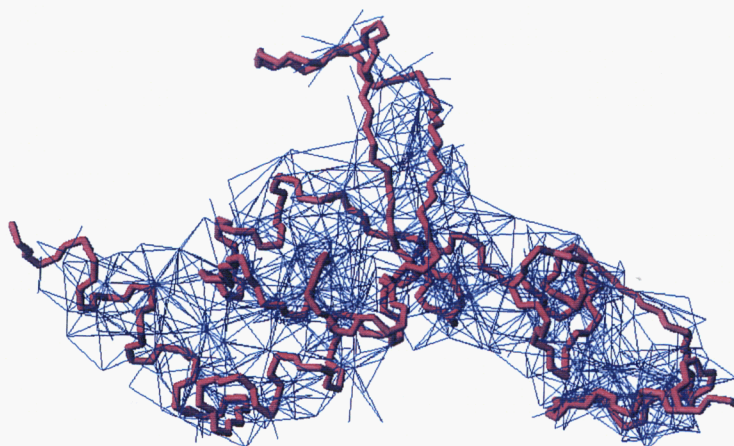


Fig. 8. The network of distance constraints used for the structure calculations of XPA₉₈₋₂₁₉.

Backbone is shown as bold magenta lines. Each of the 1,389 upper distance bounds is indicated by a thin blue line connecting the two atoms involved in the constraints.

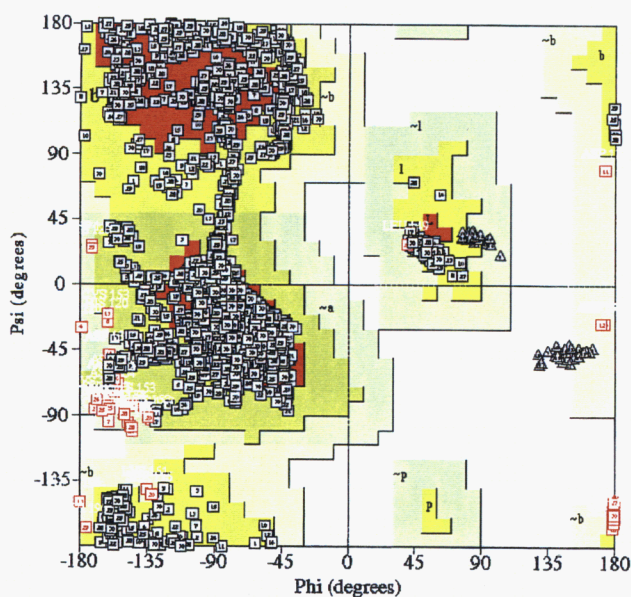
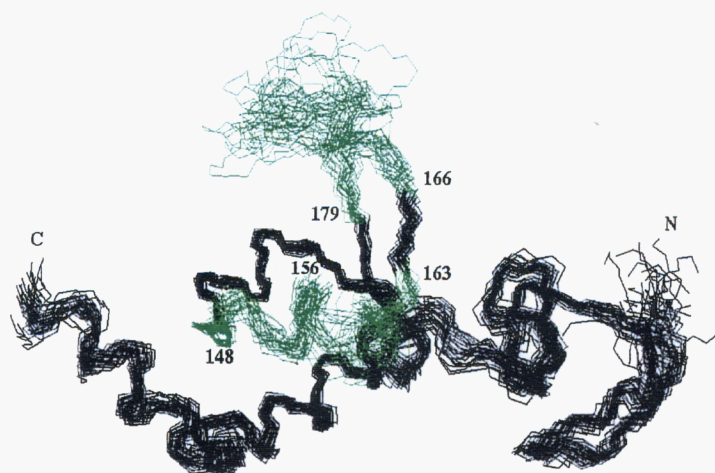


Fig. 9. Ramachandran plot of XPA₉₈₋₂₁₉.

Residues 102-155, 163-165, and 180-209 of the 30 final calculated structures were analyzed with AQUA and PROCHECK-NMR (29) software. The horizontal and vertical axes indicate the dihedral angles, ϕ (about N-C _{α} axis) and ψ (about C _{α} -C' axis), of the backbones, respectively. The numbers in the boxes indicate the molecule numbers. Most favored, additional allowed, generously allowed, and disallowed regions are shaded in black, dark grey, grey, and white, respectively. Glycine residues are shown as triangles. Statistics is listed in Table I.

(a)



(b)

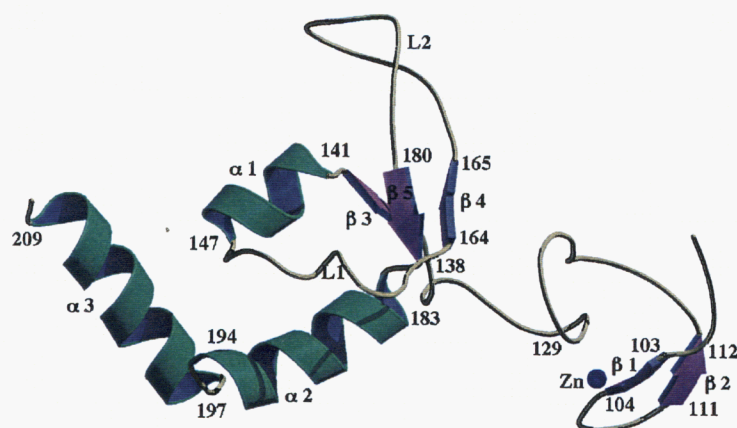


Fig. 10. Tertiary structures of the central domain of human XPA.

(a) Best-fit superposition of the backbone (N, C α , and C') atoms of the 30 final structures (residues 98 to 210), which were calculated by means of the simulated annealing procedure, and showed the lowest energy and no violation greater than 0.3 Å or 5°. The backbone atoms of residues 102 to 155, 163 to 165, and 180 to 209 are superimposed. Loops L1 (residues 148 to 163) and L2 (residues 166 to 179) are colored green. (b) Schematic ribbon drawing of the structure which shows the lowest backbone r.m.s.d. from the mean of the 30 structures, drawn with the programs, MOLSCRIPT (53) and RASTER3D (54). Secondary structure elements are indicated.

Table I. Structural statistics for the central domain of XPA.

Total number of distance constraints	1,389
Intraresidual	293
Sequential ($ i - j = 1$)	376
Medium-range ($ i - j \leq 4$)	286
Long-range ($ i - j > 4$)	381
Hydrogen bonds	47
Sulphur-sulphur	6
Number of dihedral angle constraints	83
Distance constraint violations greater than 0.3 Å	0
Dihedral angle constraint violations greater than 5°	0
R.m.s. deviations from experimental constraints	
Distance (Å)	0.0130 ± 0.0011
Angle (°)	0.325 ± 0.061
R.m.s. deviations from idealized covalent geometry	
Bonds (Å)	0.0020 ± 8.02 X 10 ⁻⁵
Angles (°)	0.487 ± 0.008
Improper (°)	0.300 ± 0.011
X-PLOR potential energy (E_{total})	171.5 ± 8.3
PROCHECK Ramachandran plot statistics	
(residues 102-209, Δ 156-162, Δ 166-179)	
Residues in the most favored regions	67.2 %
Residues in additional allowed regions	30.3 %
Residues in generously allowed regions	2.2 %
Residues in disallowed regions	0.3 %
R.m.s. deviations from the mean coordinate positions	
Zinc-containing subdomain (residues 102-129)	
Backbone heavy atoms	0.23 Å
All heavy atoms	0.75 Å
C-terminal subdomain excluding part of loop L1 and all of loop L2	
(residues 138-209, Δ 156-162, Δ 166-179)	
Backbone heavy atoms	0.52 Å
All heavy atoms	1.05 Å
Central domain excluding part of loop L1 and all of loop L2	
(residues 102-209, Δ 156-162, Δ 166-179)	
Backbone heavy atoms	0.76 Å
All heavy atoms	1.18 Å

These statistics comprise the ensemble of the final 30 simulated annealing structures from Met98 to Asn210 calculated with X-PLOR version 3.1 (27). All variances are quoted ± one SD.

Zinc-containing subdomain

The overall structure of the zinc-containing subdomain, which is composed of an antiparallel β -sheet and a helical turn, is different from those of other zinc-fingers so far reported (Fig. 11), while its local structure around the four Cys residues, including hydrogen bond networks, is common to the (Cys)₄ type zinc-fingers of GATA-1 and other zinc fingers (46). Residues 103 to 112 form a β hairpin structure which encompasses Cys105 and Cys108, whose side chains coordinate a zinc ion. The helical turn ranges from Tyr116 to Phe121, and the following loop (residues 122 to 129) provides the other two Cys residues, Cys126 and Cys129, whose side chains coordinate the zinc ion. The structure of this subdomain is dominated by the zinc ion, which is coordinated to the S γ atoms of the four Cys residues. The subdomain is also stabilized by the hydrophobic core formed by Val103, Phe112 and Met118 (Fig. 11a).

The zinc-containing subdomain has a hydrophobic patch made up of Tyr116, Leu117, Phe121, and Leu123 on the outer surface of the helical turn (Fig. 11b). These hydrophobic residues also serve as part of a hydrophobic core which contains residues in the C-terminal subdomain, Leu138, Ile165, Leu182 and Leu184. The core is formed by packing the three-stranded β -sheet of the C-terminal subdomain against the helical turn of the zinc-containing subdomain (Fig. 12). A total of 101 NOE contacts among the hydrophobic residues that form the core were observed in the NOESY spectra. Therefore, the side chain rotamers of Tyr116, Leu117, Phe121, Leu123, Leu138, and Leu182 are well defined.

The zinc-containing subdomain and the subsequent linker sequence both include many Glu and Asp residues (9/36) (Figs. 11b, 14). This highly acidic feature, which is also seen in the zinc-binding sequences of XPAs from other species, is also distinct from those of the zinc-fingers of transcription factors.

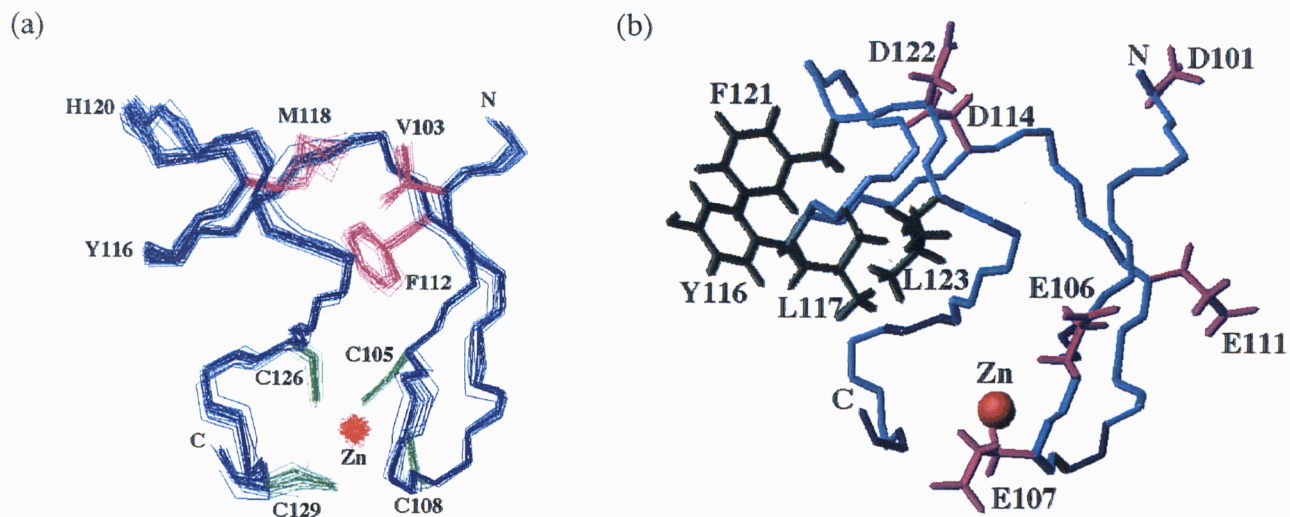


Fig. 11. Structures of the zinc-containing subdomain.

(a) Best-fit superposition of the backbone (N, C $_{\alpha}$, and C') atoms of the 30 final structures of the zinc-containing subdomain (residues 102 to 129). The backbone atoms of these residues are superimposed. The side chains of the residues in the hydrophobic core are colored red, and those of the zinc-binding Cys residues are colored green. (b) Stick representation of the zinc-containing subdomain. The locations of residues that contribute to the acidic patch in the vicinity of the β -hairpin are indicated in red. The locations of residues that contribute to the hydrophobic patch in the vicinity of the helical turn are indicated in green. Both figures were produced with the program, MOLMOL (28).

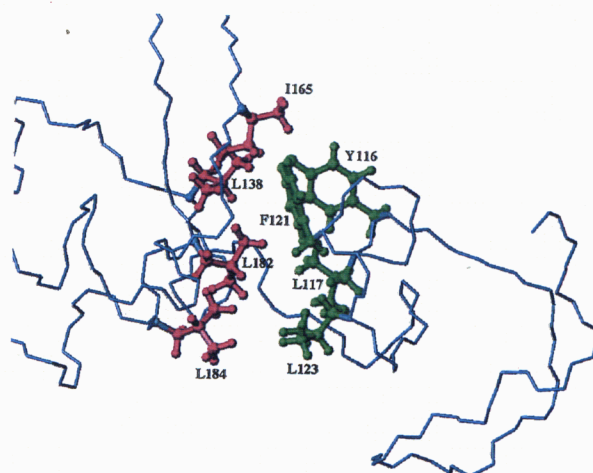


Fig. 12. The hydrophobic core formed between the zinc-containing subdomain and the C-terminal subdomain. The hydrophobic side chains of the zinc-containing subdomain that form the core are colored green, those of the C-terminal subdomain are colored red. The figure was produced with the program, MOLMOL (28).

C-terminal subdomain

The C-terminal subdomain consists of a sheet-helix-loop region (residues 138 to 182) and a helix-turn-helix region (residues 183 to 209) (Fig. 13). The sheet-helix-loop region is composed of an antiparallel β -sheet (strands β 3-5), helix α 1, and two long loops, L1 and L2. Loop L1 (residues 148 to 163) connects helix α 1 and strand β 4, and loop L2 (residues 166 to 179) connects strands β 4 and β 5. While the structure of loop L1, except for residues 156 to 162, is well defined by hydrophobic contacts with residues in helices α 1 and α 2, loop L2 apparently does not form a single definite structure based on the observation of few long-range NOEs. The signals for residues 171 to 174 could not be observed in ^{15}N - ^1H -HSQC spectra, possibly due to local exchange broadening and/or rapid exchange of the amide protons with solvent. Heteronuclear $\{^1\text{H}\}$ - ^{15}N steady-state NOE values of smaller than 0.68 at 50.7 MHz were observed for residues 157 to 162 in loop L1, for all the residues in loop L2 except residues 171 to 174 whose signals could not be observed in the HSQC spectra, and for the nine C-terminal residues of the central domain. This indicates that loop L2, part of loop L1 and the C-terminal flanking sequence are highly mobile in solution.

The helix-turn-helix region in the C-terminal subdomain consists of two long helices, α 2 (residues 183 to 194) and α 3 (residues 197 to 209), connected by the conserved Gly195-Ser196 sequence forming a short turn. This region is packed tightly against helix α 1, loop L1 and the β -sheet of the sheet-helix-loop region, stabilized by a hydrophobic core containing Ala144, Tyr148, Leu149, Leu150, Leu155, Leu162, Tyr181 and Ile186 (Fig. 13a).

The C-terminal subdomain has a large cleft between the sheet-helix-loop region and the helix-turn-helix region. The surface electrostatic potentials show that many positively charged side chains, which are contributed by well-conserved Lys141, Lys145, Lys151, Lys179, Lys204 and Arg207, are present in the cleft (Figs. 13b, 14).

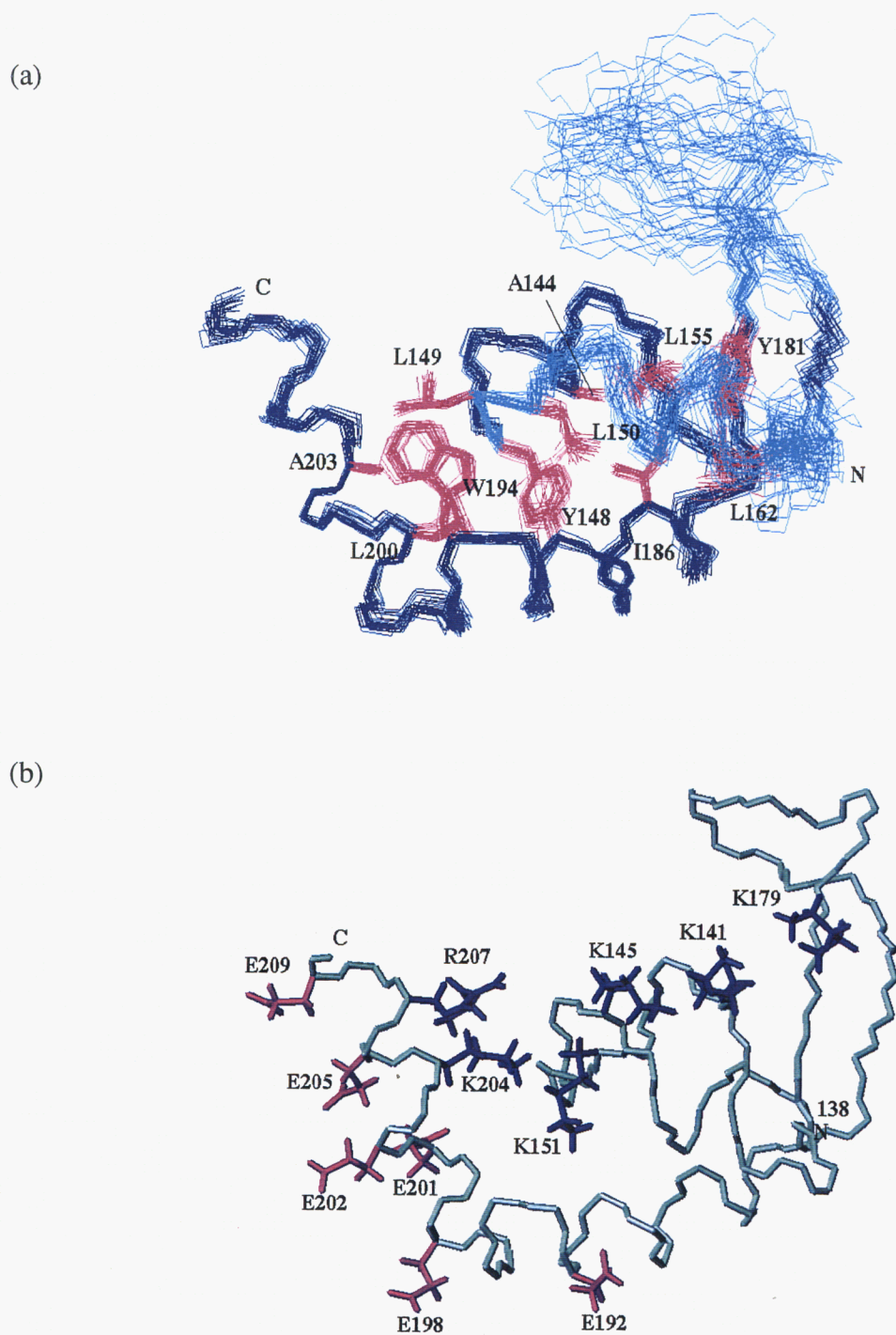


Fig. 13. Structures of the C-terminal subdomain.

(a) Best-fit superposition of the backbone (N, C_α, and C') atoms of the 30 final structures of the C-terminal subdomain (residues 138 to 209). The backbone atoms of residues 138 to 155, 163 to 165 and 180 to 209 are superimposed. Loops L1 (residues 148 to 163) and L2 (residues 166 to 179) are colored light blue. The side chains of the residues in the hydrophobic core are colored red. (b) Stick representation of the C-terminal subdomain. The well conserved residues that contribute to the positive charges in the basic cleft are colored blue. The acidic residues in helices α 2 and α 3 are colored red. Both figures were produced

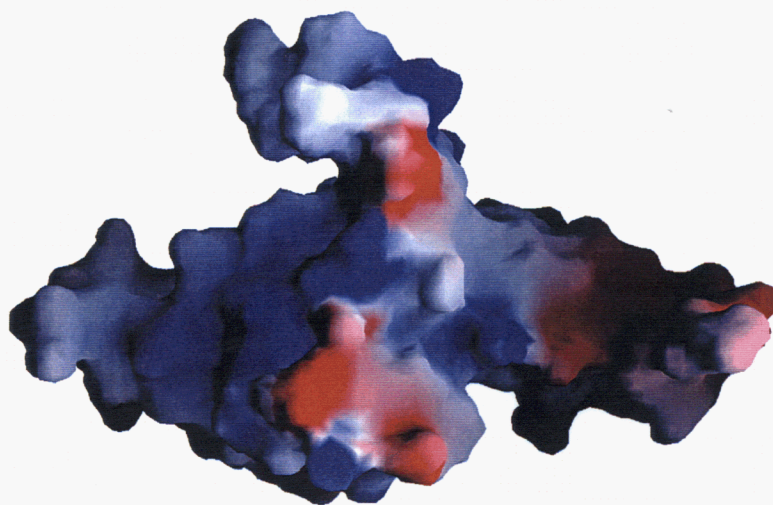


Fig. 14. Mapping of the electrostatic potential of XPA.

Distribution of the electrostatic potential displayed with GRASP (55) on the solvent-accessible surface of residues 98 to 210. Blue corresponds to positive potential and red to negative potential. The presence of a positively charged cleft is evident in the C-terminal subdomain. In the zinc-containing subdomain negatively charged patches are dominant. The molecular orientation is the same as in Fig. 16.

● DNA and RPA70 Binding Surfaces

Interaction with DNA

For the identification of the DNA binding surface of the central domain of XPA, chemical shift perturbation experiments were done with ^{15}N -labelled XPA₉₈₋₂₁₉. Selective chemical shift perturbation and/or broadening were observed for the signals in the ^{15}N - ^1H -HSQC spectrum after mixing with an equimolar amount of the 24-mer oligonucleotide treated with cisplatin. This chemotherapeutic agent reacts with the oligonucleotide to form a single 1,3-intrastrand d(GpTpG)-cisplatin crosslink (58). Almost all of the signals that showed remarkable chemical shift

perturbation (defined as more than 0.08 ppm for ^{15}N or 0.02 ppm for ^1H chemical shifts) or broadening (defined as peak intensities decreased to $< 50\%$ of the original values) were attributed to the amide residues in the basic cleft of the C-terminal subdomain (Figs. 15a, 15c, 14, and 16). On the other hand, smaller and more uniform signal losses with no detectable chemical shift change were observed for the residues in the zinc-containing subdomain. The peak intensities for the 22 well-resolved signals in the subdomain became 61.3 ± 7.2 (1σ) % of the original values. These small and uniform losses are caused mainly by an increase of the effective molecular mass upon formation of the XPA-DNA complex. These findings suggest that the basic cleft and the surrounding region are involved in the interaction with DNA, but that the zinc-containing subdomain is not.

Interaction with RPA70

The interaction of the central domain of XPA with RPA70₁₈₁₋₄₂₂ was also examined by chemical shift perturbation experiments with ^{15}N -labelled XPA₉₈₋₂₁₉. Selective signal losses were observed in the ^{15}N - ^1H -HSQC spectrum after mixing with an equimolar amount of RPA70₁₈₁₋₄₂₂. Signal losses are observed when the chemical shifts are perturbed considerably by binding, the exchange rate between the free and bound states is comparable to the chemical shift difference between the two states, or the ligand binding enhances the amide proton exchange rate with a solvent. Almost all of the signals for which the peak intensities decreased to $< 65\%$ of the original values were attributed to the residues in the zinc-containing subdomain (Figs. 15b, 15c, 14, and 16), suggesting that this subdomain serves as an RPA70-binding surface. Smaller and more uniform signal losses were observed for the residues of the C-terminal subdomain. The peak intensities for the 30 well-resolved signals in the subdomain became 73.6 ± 5.2 (1σ) % of the original values. These small and uniform losses are due mainly to an increase of the effective molecular mass upon formation of the XPA-RPA complex.

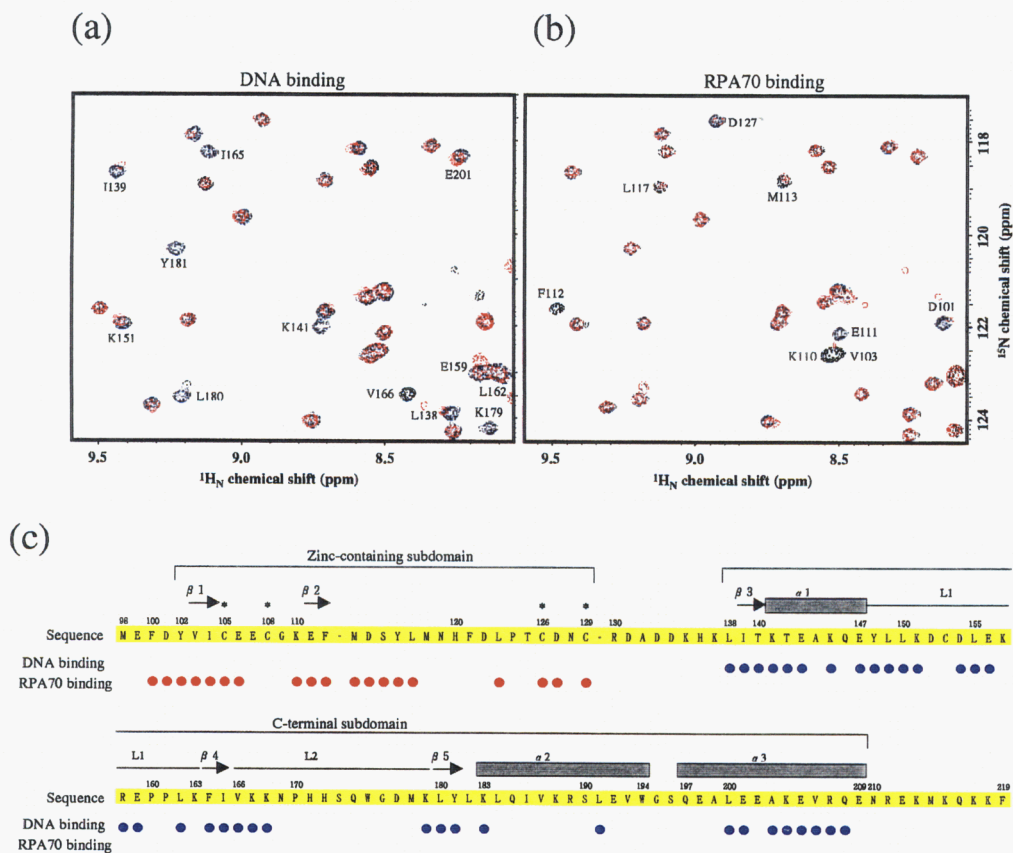


Fig. 15. Identification of the binding surfaces for DNA and RPA70.

(a) ^{15}N - ^1H -HSQC spectra of the ^{15}N -labelled XPA₉₈₋₂₁₉ and the complex with the 24-mer oligonucleotide treated with cisplatin. The reference spectrum of the 50 μM protein with a ^1H frequency of 800.13 MHz is shown in black. The assignments of the amide resonances are indicated. The spectrum of the protein in the presence of a molar equivalent of the chemically damaged oligonucleotide is shown in red. (b) ^{15}N - ^1H -HSQC spectra of ^{15}N -labelled XPA₉₈₋₂₁₉ and the complex with RPA70₁₈₁₋₄₂₂. The reference spectrum of the 50 μM protein with a ^1H frequency of 800.13 MHz is shown in black. The spectrum of the protein in the presence of a molar equivalent of RPA70₁₈₁₋₄₂₂ is shown in red. (c) Residues of which the amide signals showed chemical shift perturbation and/or broadening upon binding to the oligonucleotide treated with cisplatin or to RPA70₁₈₁₋₄₂₂ are indicated by blue and red dots under the primary sequence of the central domain of XPA, respectively. The asterisks on the sequence indicate the zinc-coordinated Cys residues.

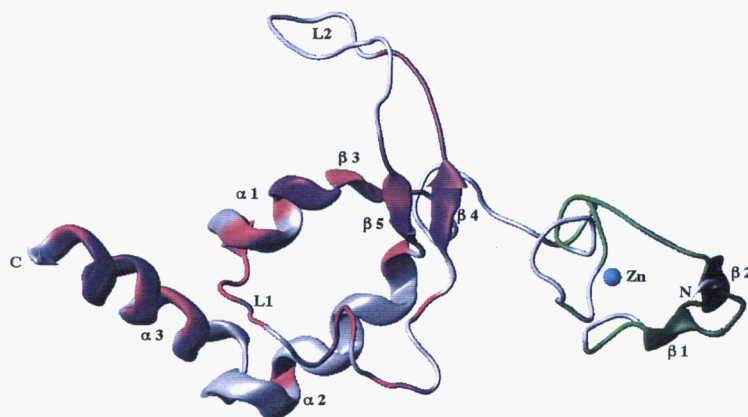


Fig. 16. Mapping of the interaction surfaces of XPA.

Mapping of the XPA residues with chemical shift perturbation or broadening effects in the ^{15}N - ^1H -HSQC spectra. The residues of which the amide resonances were perturbed upon complex formation with the cisplatin-damaged 24-mer oligonucleotide are indicated in magenta (chemical shift perturbation defined as > 0.08 ppm for ^{15}N or > 0.02 ppm for ^1H , or broadening defined as peak intensities decreased to $< 50\%$ of the original values), and the residues of which the amide resonances showed specific broadening upon complex formation with RPA70₁₈₁₋₄₂₂ are colored green (the peak intensities decreased to $< 65\%$ of the original values).

● Dynamics

^{15}N relaxation data

The relaxation data, ^{15}N T_1 , T_2 , $\{^1\text{H}\}$ - ^{15}N steady-state NOE, and T_1/T_2 ratios, were obtained by analysis of ^1H -detected ^{15}N - ^1H correlation spectra of XPA₉₈₋₂₁₉ at 50.7 MHz and 81.1 MHz ^{15}N frequency fields (Fig. 17). Most data for the ^{15}N spins in the five N-terminal residues (98 to 102), the five residues (170 to 174) in loop L2, and the four C-terminal residues (216 to 219) could not be obtained due to signal broadening or severe signal overlapping in the spectra. Smaller $\{^1\text{H}\}$ - ^{15}N NOE values (< 0.55 at 50.7 MHz) were found for a part of loop L1 (residue 158), most of loop L2 (residues 168 to 179, except for non-detectable resonances), and the latter part of helix $\alpha 3$ (residues 206 to 209), and markedly smaller or negative $\{^1\text{H}\}$ - ^{15}N NOE values (< 0.2 at 50.7 MHz) were found for the N-terminal region (residue 102) and the C-terminal region (residues 212 to 218, except for non-detectable resonances). The T_1/T_2 ratios showed wide distribution ranges ($10.8 \pm 2.2(1\sigma)$ at 50.7 MHz and $25.9 \pm 7.7(1\sigma)$ at 81.1 MHz), which implied the presence of anisotropy in the rotational diffusion of the domain.

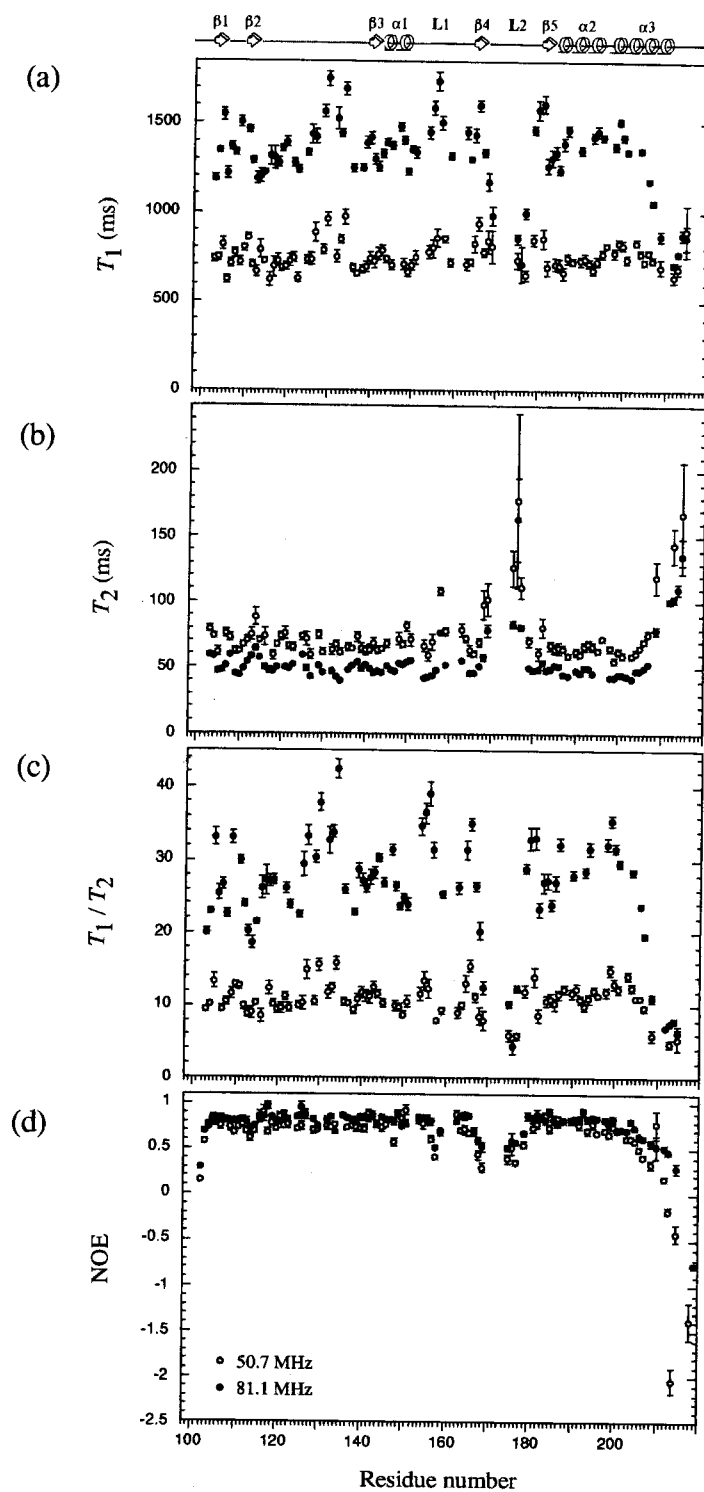


Fig. 17. Plots of amide ^{15}N T_1 , T_2 , T_1/T_2 , and NOE against the residue number.

The data were measured at ^{15}N frequencies of 50.7 MHz (open circles) and 81.1 MHz (filled circles). (a) Longitudinal relaxation times, T_1 . (b) Transverse relaxation times, T_2 . (c) T_1/T_2 ratios. (d) Heteronuclear $\{^1\text{H}\}$ - ^{15}N steady-state NOE values defined as I/I_0 , where I and I_0 are the intensities of the peaks with and without ^1H saturation, respectively. For plots a-d, the error bars indicate the $\pm 1\sigma$ values of the uncertainties of the data. The secondary structures are indicated at the top.

J(0) analysis

The effective spectral density function at zero-frequency, $J_{\text{eff}}(0)$, is defined as the sum of $J(0)$ and λR_{ex} (35). Because R_{ex} values are proportional to the square of the magnetic field strength, and $J(0)$ values are independent of the field strength, R_{ex} values can be determined from $J_{\text{eff}}(0)$ values obtained at two different field strengths. Fig. 18 shows the $J_{\text{eff}}(0)$ values at 50.7 MHz and 81.1 MHz, and the calculated R_{ex} values. The mean and SD of the R_{ex} values at 50.7 MHz was $1.54 \pm 0.97 \text{ s}^{-1}$. High R_{ex} values ($> 3.5 \text{ s}^{-1}$ at 50.7 MHz) were found for residues 133 and 134 in the linker sequence, and residues 154 and 156 in loop L1.

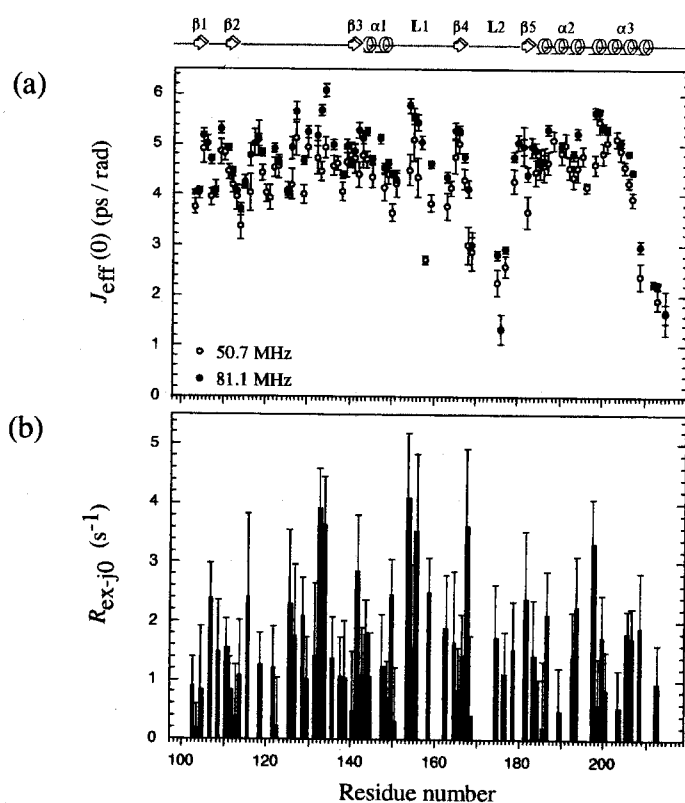


Fig. 18. Plots of the results of $J(0)$ analysis against the residue number.

(a) Reduced effective spectral density functions at zero frequency, $J_{\text{eff}}(0)$, estimated at ^{15}N frequencies of 50.7 MHz (open circles) and 81.1 MHz (filled circles). (b) Chemical exchange rates, $R_{\text{ex-j0}}$, were calculated from the $J_{\text{eff}}(0)$ values estimated at the two static magnetic fields. For plots a and b, the error bars indicate the $\pm \text{SD}$ (1σ) values of the uncertainties of the data. The secondary structures are indicated at the top.

Isotropic, axial symmetric, fully anisotropic models

To determine which diffusion model is more appropriate for analysis of the relaxation data, the experimental T_1/T_2 ratios were fitted to each of the isotropic, axially symmetric, and fully anisotropic diffusion models. The results of the fitting were evaluated by means of F-statistical tests (40). The T_1/T_2 ratios at 50.7 MHz were used for the analysis since they were expected to involve less contribution from the chemical exchange rates, R_{ex} , than those at 81.1 MHz. The T_1/T_2 ratios of the residues whose $\{^1H\}$ - ^{15}N NOE values were smaller than 0.65 at 50.7 MHz, or whose R_{ex} values were greater than 2.5 s^{-1} on $J(0)$ analysis were excluded from the analysis. χ^2 values, 320.0, 194.3 and 182.9, were obtained for the isotropic, axially symmetric, and fully anisotropic diffusion models, respectively. For these χ^2 values, the F-statistical values were calculated to be 11.9 between the isotropic and axially symmetric models, and 1.7 between the axially symmetric and fully anisotropic models. These F-statistical values were used to determine whether the improvement in the fitting, expressed as the χ^2 values, by using a more complicated model is significant or merely arises due to the incorporation of additional parameters that causes reduction of the degree of freedom (33). The F-statistical value for the isotropic and axially symmetric models (11.9) was significantly larger than the corresponding tabulated 95% critical value of 2.8 obtained from the statistical table presented by Bevington *et al.* (34). On the other hand, that for the axially symmetric and fully anisotropic models (1.7) was smaller than the corresponding tabulated 95% critical value of 3.2. These results show that the improvement in the fitting using the axially symmetric diffusion model instead of the isotropic model was statistically significant, but that using the fully anisotropic model instead of the axially symmetric model was merely due to the incorporation of additional parameters.

Another test for the validity of the axially symmetric model was performed using in-house programs. First, the T_1/T_2 ratios at 50.7 MHz were randomly assigned to the 59 backbone amide bond vectors in XPA. This random assignment removes the correlation between the orientations of the amide bond vectors and the T_1/T_2 ratios (40). Then, an axially symmetric model was used for the fitting to these randomly assigned T_1/T_2 ratios. The procedures for the random assignment and the fitting to these data were independently repeated 200 times, and then the statistical significance was evaluated. The average and minimum values of χ^2 obtained through the 200 simulations were 299.8 and 242.1, respectively. As the value of χ^2 , 194.3, for the fitting based on the correct amide bond vectors was smaller than the minimum χ^2 value of 242.1 in the 200 random simulations, it was shown that the correlation between the T_1/T_2 values and the orientations of the

amide bond vectors was statistically significant.

Model-free analyses

The relaxation data were analyzed by means of model-free formalism under the assumption of isotropic rotational diffusion, or the assumption of axially symmetric anisotropic rotational diffusion. The results were compared for the judgement on which model was more appropriate for the data i.e., the isotropic diffusion model or the axially symmetric anisotropic diffusion model.

Initial estimation of τ_m and D

The overall rotational correlation time, τ_m , estimated from the mean (10.81 ± 0.65 [uncertainty]) of the 61 one σ trimmed T_1/T_2 ratios at 50.7 MHz was 12.13 ± 0.43 ns. The number of spins for which the $\{^1\text{H}\}$ - ^{15}N NOE values were smaller than 0.65 at 50.7 MHz was 23, and the number of spins for which the R_{ex} values estimated on $J(0)$ analysis were larger than 2.5 s^{-1} at 50.7 MHz was seven. These spins were excluded from the calculation of the axially symmetric rotational diffusion tensor, D . The remaining 59 T_1/T_2 ratios at 50.7 MHz and the coordinates of XPA (PDB: 1xpa) were used for estimation of the principal values of the tensor. The ratio of the diffusion constants parallel and perpendicular to the long axis of the symmetric rotor (D_{\parallel}/D_{\perp}) was 1.38 ± 0.04 , and the average correlation time ($1/[2D_{\parallel} + 4D_{\perp}]$) was 12.03 ± 0.05 ns (Fig. 19). The orientations of the unique axes of the calculated diffusion tensor and the inertia moment estimated only from the molecular coordinates agreed well (7.9° difference).

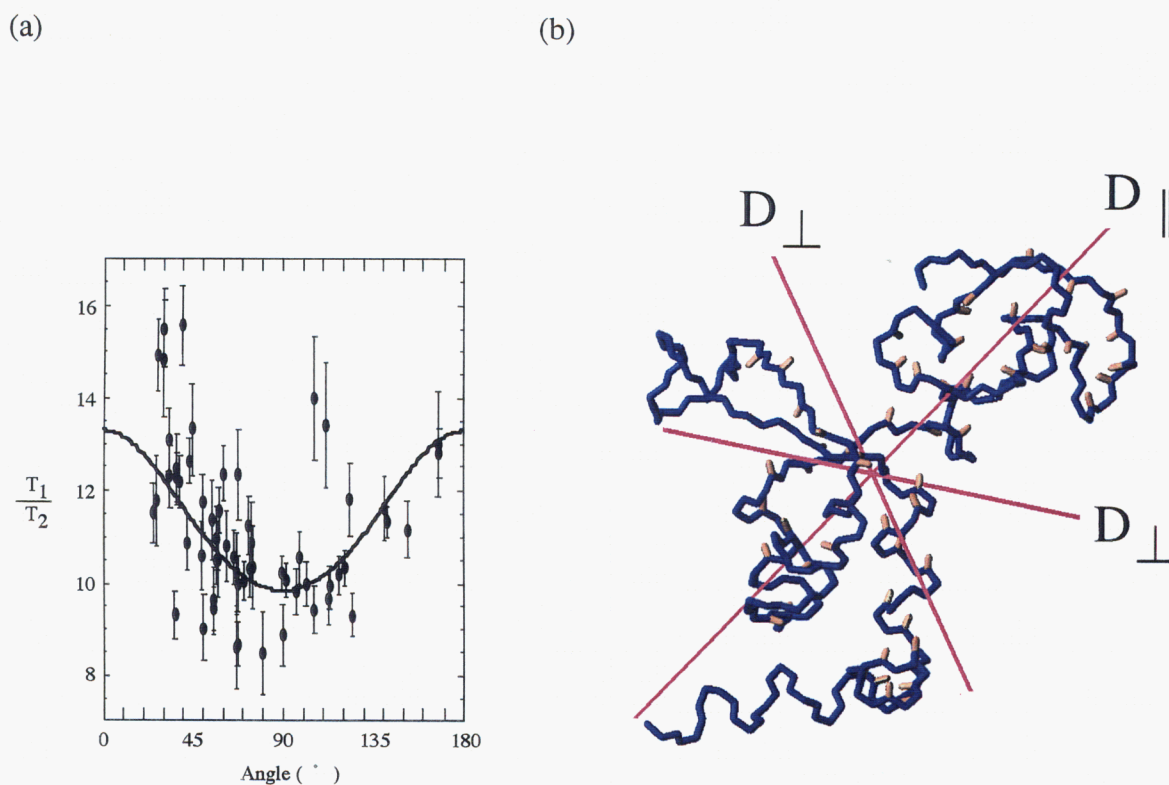


Fig. 19. Axially symmetric anisotropic character of the rotational diffusion of the central domain of XPA.

(a) Plots of the observed ^{15}N T_1/T_2 ratios at 50.7 MHz against the angle, θ , between the NH bond vectors and the unique axis of the diffusion tensor (D_{\parallel}) for the coordinates of XPA (PDB: 1xpa). The ^{15}N T_1/T_2 ratios of the spins which showed significant internal motions on a time scale of longer than a few hundred picoseconds ($\{^1\text{H}\}$ - ^{15}N NOE < 0.65 at 50.7 MHz) or conformational exchange (R_{ex} estimated on $J(0)$ analysis at 50.7 MHz $> 2.5 \text{ s}^{-1}$) were not used for determination of the rotational diffusion tensor, D , and are not plotted in a. The curved line represents the theoretical dependence of the T_1/T_2 ratios on the angle, θ , for the determined diffusion tensor with an anisotropy (D_{\parallel}/D_{\perp}) of 1.38 and an average correlation time ($1/[2D_{\parallel} + 4D_{\perp}]$) of 12.03 ns. The error bars indicate the $\pm \text{SD}$ (1σ) values of the uncertainties of the T_1/T_2 ratios. (b) Stick representation of the backbone and NH bonds of the central domain of XPA (PDB: 1xpa) with the principal axis orientations of the determined rotational diffusion tensor, D . Only the NH bonds whose coordinates were used for determination of the tensor are drawn. The solid lines labelled D_{\parallel} and D_{\perp} correspond to the respective orientations of the tensor. The ratio of the lengths of the lines labelled D_{\parallel} and D_{\perp} equals the actual anisotropy of the tensor ($D_{\parallel}/D_{\perp} = 1.38$). The orientations of D_{\perp} can be rotated arbitrarily about the unique axis (D_{\parallel}) since the tensor, D , is assumed to be axially symmetric.

Selection of the dynamic models

The experimental ^{15}N T_1 , T_2 , and NOE values were fitted to the dynamic models 1 to 5 with the isotropic diffusion model with τ_m fixed at the initially estimated value, 12.13 ns, using a statistical model selection protocol (33). Likewise, they were fitted to the dynamic models 1 to 4 with the axially symmetric diffusion model with a fixed orientation and magnitude of the diffusion tensor, D . Table II summarizes the number of spins for which each dynamic model was selected. One of the two simpler models, 1 or 2, was chosen for 15 spins with the isotropic model, while one of them was chosen for 28 spins with the anisotropic model, suggesting that the anisotropic diffusion model is more appropriate than the isotropic model for model-free analysis.

Table II. The number of residues selected for each dynamic model on model-free analyses with isotropic and anisotropic rotational diffusion models.

classification ^a	model 1	model 2	model 3	model 4	model 5	not fit	total
parameters ^b	S^2	S^2, τ_{ef}	S^2, R_{ex}	S^2, τ_{ef}, R_{ex}	S_f^2, S^2, τ_{es}		
isotropic ^c	3 ^e ,2 ^f ,9 ^g (14) ^h	1,0,0(1)	2,3,9(14)	4,0,15(19)	3,3,21(27)	0,0,15(15)	13,8,69(90)
anisotropic ^d	5,4,14(23)	4,0,1(5)	0,2,8(10)	3,0,21(24)	none	1,1,23(25)	13,7,67(87)

^a The five dynamic models applied for the fitting in model-free analyses developed by Lipari and Szabo (37, 38). ^b The parameters optimized for each dynamic model. ^c Isotropic rotational diffusion was assumed. ^d Axially symmetric anisotropic rotational diffusion was assumed. ^e The number of residues for which only the data (T_1 , T_2 , and NOE) at the ^{15}N frequency of 50.7 MHz were available and fitted the corresponding dynamic model. ^f The number of residues for which only the data at the ^{15}N frequency of 81.1 MHz were available and fitted the dynamic model. ^g The number of residues for which the data at the ^{15}N frequencies of both 50.7 MHz and 81.1 MHz were available and fitted simultaneously the dynamic model. ^h The sum of the numbers (e), (f) and (g).

Optimization of τ_m , D , and internal motion parameters

The overall rotational correlation time, τ_m , for the isotropic diffusion model or the diffusion tensor, D , for the anisotropic diffusion model, and the internal motion parameters were simultaneously optimized with the selected dynamic model for each spin. The optimized overall correlation time, τ_m , was 12.23 ns, the average correlation time of the diffusion tensor ($1/[2D_{\parallel} + 4D_{\perp}]$) was 11.96 ns, and the ratio of the principal values of the diffusion tensor (D_{\parallel}/D_{\perp}) was 1.39. The associated internal motion parameters are shown in Figs. 20, 21 and 22. The overall correlation time, τ_m , and the average correlation time derived from the diffusion tensor, D , showed good agreement (0.27 ns difference). The orientation of the unique axis of the diffusion tensor determined on model-free analysis is close to that of the diffusion tensor determined from the T_1/T_2 ratios (6.1° difference), and also close to that of the inertia moment estimated from the molecular coordinates (2.6° difference). The S^2 , R_{ex} , and τ_{ef} values determined with the isotropic and anisotropic models did not exhibit statistically significant differences, as described under Discussion.

Relatively smaller S^2 values (< 0.75) were found for residues 130 and 134 in the linker sequence. Significantly smaller S^2 values (< 0.6) for the isotropic model were found in parts of loop L1 (residue 158), loop L2 (residues 169, 176 and 177), and the C-terminal region (residues 209, 212, 213 and 215).

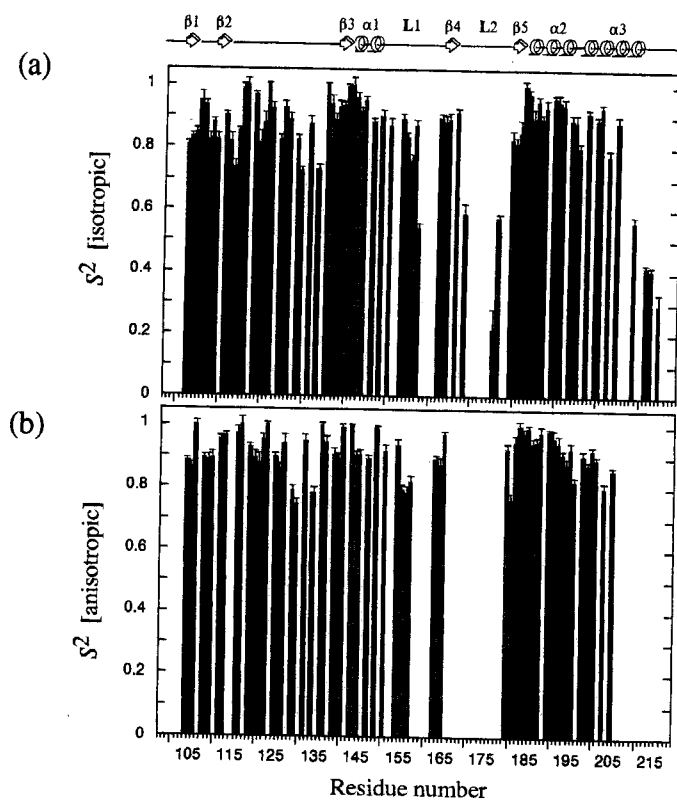


Fig. 20. Plots of order parameters, S^2 , against the residue number.

The S^2 values were determined on model-free analyses with the isotropic rotational diffusion model (a), and the axially symmetric anisotropic rotational diffusion model (b). For plots a and b, the error bars indicate the $\pm SD(1\sigma)$ values of the uncertainties estimated by Monte Carlo simulations. The secondary structures are indicated at the top.

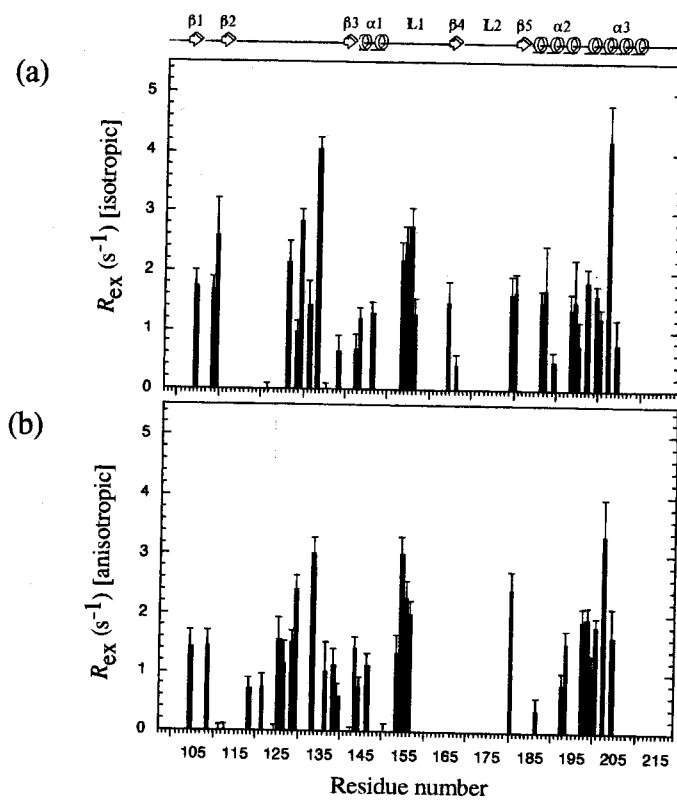


Fig. 21. Plots of chemical exchange rates, R_{ex} , against the residue number.

The R_{ex} values were determined on model-free analyses with the isotropic rotational diffusion model (a), and the axially symmetric anisotropic rotational diffusion model (b). For plots a and b, the error bars indicate the $\pm SD(1\sigma)$ values of the uncertainties estimated by Monte Carlo simulations. The secondary structures are indicated at the top.

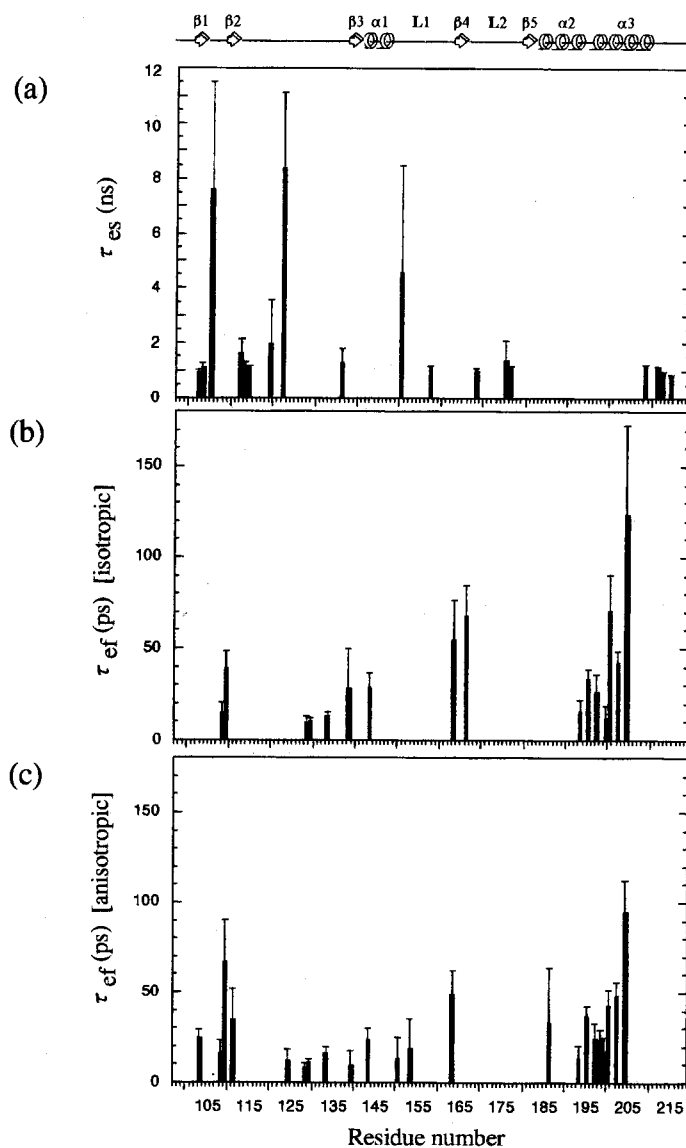


Fig. 22. Plots of effective correlation times, τ_{ef} and τ_{es} , against the residue number.

(a) The slow internal correlation times, τ_{es} , were determined on model-free analysis with the isotropic rotational diffusion model using the extended formula for the spectral density function proposed by Clore *et al.* (39). This extended formula was not used for model-free analysis with the axially symmetric anisotropic rotational diffusion model. The fast internal correlation times, τ_{ef} , were determined on model-free analyses with the isotropic rotational diffusion model (b), and the axially symmetric anisotropic rotational diffusion model (c). For plots a-c, the error bars indicate the $\pm SD$ (1σ) values of the uncertainties estimated by Monte Carlo simulations. The secondary structures are indicated at the top.

DISCUSSION

● Description of the Structure

Comparison with other structures

Comparison of the amino acid sequence of the central domain of human XPA with those of XPAs from other eukaryotes, including the yeast homologue of XPA, RAD14, indicates that the level of sequence identity ranges from 28% (yeast) to 93% (mouse). The similarity of the tertiary structure of XPA to those of known proteins was analyzed with DALI server version 2.0 (43). It showed that there is no previously determined structure with a Z score of > 1.9 in the database.

In the zinc-containing subdomain of XPA, a series of hydrogen bond networks has been found around the four Cys residues: Cys105 NH - Lys110 O, Glu107 NH - Cys105 S_γ, Cys108 NH - Cys105 S_γ, Gly109 NH - Cys105 O, Lys110 NH - Cys108 S_γ, Asn128 NH - Cys126 S_γ, Cys129 NH - Cys126 S_γ, Arg130 NH - Cys126 O, and Cys126 NH - Cys105 S_γ. The presence of these hydrogen bonds was indicated by both the presence of slow exchanging amide protons and structures calculated without these hydrogen bond constraints. These local structure elements found in XPA are common in the (Cys)₄ type zinc-fingers of erythroid transcription factor GATA-1 (44) and the glucocorticoid receptor (GRE) (45, 46). However, alignment of the zinc-binding sequence of XPA, Cys-X-X-Cys-(X)₁₇-Cys-X-X-Cys, with that of GATA-1 showed that only the positions of four Cys residues and a Pro residue (XPA position 124) are identical. Moreover, the zinc-finger of GATA-1 has more basic residues than acidic residues, as usually seen in the zinc-fingers of the DNA binding domains of transcription factors, whereas the zinc-containing subdomain of XPA is highly acidic (Fig. 14).

Exons/introns and tertiary structure elements

Satokata *et al.* have characterized the genomic structure of the human *xpa* gene, and identified six exons in the gene (47). Interestingly, the genomic structure of the *xpa* gene shows quite good correlation to the tertiary structure elements (Fig. 23). Exon 3 (residues 96 to 130) encodes the entire zinc-containing subdomain, exon 4 (residues 131 to 185) encodes the linker (residues 130 to 137) and the sheet-helix-loop region (residues 138 to 182), and exon 5 (residues 186 to 224) encodes most of the helix-turn-helix region (residues 183 to 209) and the C-terminal flanking sequence (residues 211 to 219). This result indicates that the introns are located at sites

corresponding to the inter-module junctions of the central domain of XPA. The correlation between the functional domains of XPA and its genomic structure does not seem to be limited to the central domain. Previous analysis involving truncated XPA has shown that a particular region, residues 59 to 97, is necessary for ERCC1 binding, and this region almost corresponds to the region encoded by exon 2 (residues 58 to 95) (11, 14-16). The TFIIH binding site, residues 226 to 273, is encoded by exon 6 (residues 225 to 273) (12). Further structure determination will reveal the relationship between the exons/introns and the tertiary structure elements of these regions (48, 49).

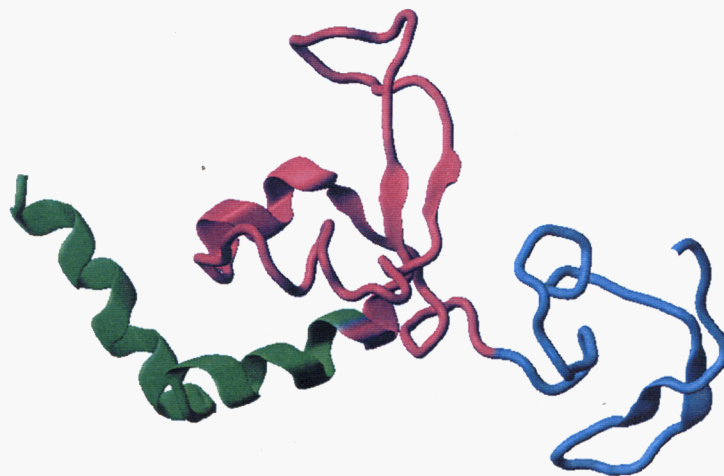


Fig. 23. Relationship between the genomic structure and tertiary structure elements of XPA. Exons 3, 4, and 5 are drawn in cyan, magenta, and green, respectively.

● Interactions of XPA with DNA and RPA70

Interaction with damaged DNA

The results of the chemical shift perturbation experiment on mixing with damaged DNA suggest that the basic cleft and the surrounding region are involved in the interaction with DNA, but that the zinc-containing subdomain does not directly interact with DNA. Visual inspection revealed that the internal curvature of the basic cleft fits the diameter of standard B-form double-stranded DNA (Fig. 24b). All the signals attributed to the positively charged residues in the cleft showed induced chemical shift changes upon the addition of DNA, suggesting that they might interact ionically with the backbone phosphate groups of DNA.

The induced chemical shift perturbation was most remarkable for the C-terminal end of helix $\alpha 3$ (^{15}N chemical shift changes > 0.3 ppm or peak intensity decreases $< 35\%$ for residues 204 to 208). The importance of helix $\alpha 3$ can be seen in the following observations. XP129 is an UV-resistant revertant of an XP group A cell line (XP12RO) that is homozygous for the nonsense mutation at amino acid position 207. In XP129 this stop codon has been mutated further and encodes the Gly residue in one allele instead of the wild type Arg residue. XP129 can repair (6-4) photoproducts normally but not cyclobutane dimers (CPD), while XP12RO cannot repair either kind of damage (59, 60). It was also shown that the rate of genome overall repair of both CPD and (6-4) photoproducts in the transfectant XP12ROSV cells overexpressing R207G mutant was indistinguishable from that in the transfectant overexpressing wild type XPA, whereas no detectable repair of CPD and (6-4) photoproducts was observed in control XP12ROSV cells (79). These results suggest that the C-terminus of helix $\alpha 3$ is important for the recognition of DNA damages and/or interaction with other NER factors. In addition, the truncated XPA₄₉₋₁₉₄, which lacks the whole $\alpha 3$ helix, lost its DNA binding activity (8).

Specificity to damaged DNA

A similar pattern of chemical shift perturbation was observed when complex was formed with the same 24-mer oligonucleotide with no damage. This similarity suggests that the same binding surface is shared by both undamaged and damaged DNAs.

It has been shown that XPA and XPA-RPA bind with moderately higher affinity to damaged DNA compared with undamaged DNA by various methods including filter binding and gel retardation assays (6, 8, 9, 10). For example, the filter binding experiments done by Kuraoka *et al.* showed the affinities of the central domain of XPA for 2,686 bp DNAs with and without

multiple damages (8). According to our calculation based on these results, the dissociation constant, K_d , for XPA₉₈₋₂₁₉ and DNA damaged by UV or cisplatin is approximately 4×10^{-8} M, and the K_d for XPA₉₈₋₂₁₉ and the undamaged DNA is in the range of 3×10^{-6} M to 6×10^{-7} M. However, in these experiments done so far, large DNA fragments with random damages were used for detecting DNA-protein complexes. There have been no reports on the preferential binding to a small DNA fragment with a single lesion like the one-site platinated 24-mer oligonucleotides used in this experiment.

No large difference between the chemical shift perturbations with damaged and undamaged DNA is possibly due to the low binding specificity of XPA alone to damaged over undamaged DNA (8). Hence, multi-site damages on larger DNA may be necessary for the detection of the low binding specificity. The moderate specificity of XPA or XPA-RPA to damaged DNA is not enough to account for the capacity of the human NER system to repair rare damages within a huge amount of human DNA. In addition, it has been recently shown that a stable complex of high specificity can be detected only when XPA-RPA, TFIIF, XPC-HHR23B, XPG and ATP are present *in vitro* (18). Therefore, it is possible to assume that XPA and/or RPA at first find damaged sites on DNA, and then additional several NER factors together with XPA/RPA perfectly recognize damaged sites by making the complex more stable.

Substrates for human NER cover a wide variety of damages on DNA. There is also a report showing that more affinity was found for single-stranded DNA over double-stranded DNA (6). Thus, specific chemical groups on DNA might not be determinants for the recognition of the damage. NER factors may recognize an abnormal DNA structure with a single-stranded character as a result of destabilization or distortion by damages. Under this assumption, the 24-mer oligonucleotide with one cisplatin damaged site which was used in this experiment may be too short to exhibit a single-stranded character as to attract XPA preferentially.

Interaction with RPA70

RPA is a heterotrimeric single-stranded-DNA-binding protein, which is involved in DNA replication, homologous recombination, and NER, and its interaction with XPA has been shown to be essential for NER (9, 10, 61, 62). Although XPA binds to the two largest subunits of RPA, RPA70 and RPA34, the deletion analyses of XPA showed that RPA70 contributes more predominately to the XPA-binding than RPA34 does (9-11). By means of limited proteolysis and mutational analysis, and from its crystal structure, Pfuetzner *et al.* have shown that residues 181 to 422 of RPA70, RPA70₁₈₁₋₄₂₂, form a compact structural domain that exhibits single-stranded

DNA binding activity (52, 63).

The results of the chemical shift perturbation experiment on mixing with RPA70₁₈₁₋₄₂₂ suggest that the zinc-containing subdomain serves as an RPA70-binding surface of XPA. All of the amide residues with peak intensities decreased to < 20% of the original values are located on the β hairpin and its vicinity (residues 101 to 114). This region contains five acidic residues. The binding between the central domain of XPA and RPA70₁₈₁₋₄₂₂ is sensitive to ionic strength. As the KCl concentration in the sample solution was increased from 28 to 77 mM, the signal loss of XPA₉₈₋₂₁₉ caused by its binding to RPA70₁₈₁₋₄₂₂ was much reduced. These results suggest that the molecules may interact through the electrostatic force between the negatively charged patch around the β hairpin of XPA and a positively charged patch in RPA70₁₈₁₋₄₂₂.

Through deletion analyses, Saijo *et al.* have shown that residues 98 to 187 of XPA, which include the zinc-containing subdomain and the sheet-helix-loop region (residues 138 to 182) of the C-terminal subdomain, are sufficient for binding to RPA70, but further deletion of the sheet-helix-loop region would cause the loss of RPA70 binding (11). These results indicate that the sheet-helix-loop region of the C-terminal subdomain is also required for RPA70 binding, although the present data have shown that the C-terminal subdomain does not directly interact with RPA70₁₈₁₋₄₂₂. The sheet-helix-loop region and its vicinity contain Leu138, Ile165, Leu182, and Leu184, whose side chain methyl groups make hydrophobic contacts with the hydrophobic patch formed by Tyr116, Leu117, Phe121, and Leu123 in the zinc-containing subdomain (Fig. 12). These contacts can be necessary to stabilize the fine structure of the zinc-containing subdomain as well as the conformation of the C-terminal subdomain. There is also a possibility that the sheet-helix-loop region of XPA serves as another binding surface for the N- or C-terminal region, outside of RPA70₁₈₁₋₄₂₂, of RPA70.

● Protein Dynamics

Selection of the diffusion models

The central domain of XPA consists of a zinc-containing subdomain, a C-terminal subdomain, and a linker sequence which connects these two subdomains. The relative magnitudes of the inertia moment calculated from the molecular coordinates (PDB: 1xpa) were 1.00 : 0.90 : 0.41, indicating that the overall shape of the molecule is rather elliptical. This is coincident with the observation that the T_1/T_2 ratios exhibit wide dispersion, which is mainly due to the anisotropy of the rotational diffusion (Fig. 17c). These observations suggest that an anisotropic diffusion model is preferable for the analysis of the ^{15}N relaxation data.

The F-statistical and random assignment tests have shown that the axially symmetric anisotropic model is more appropriate than the isotropic model for the analysis of the ^{15}N relaxation data. However, the relaxation data were analyzed with the axially symmetric model, as well as with the isotropic model for the following two reasons. First, the analysis with an axially symmetric model for the ^{15}N spins requires information on the orientations of the corresponding N-H bond vectors. However, they are not available for flexible parts of the molecule, for example, several residues in loop L2. Second, it is generally difficult to distinguish the contribution of overall motional anisotropy from that of slow internal motions (50).

Comparison of the dynamic parameters

The order parameters, S^2 , determined on model-free analyses with the isotropic and axially symmetric anisotropic rotational diffusion models were compared (Fig. 20). The r.m.s. of the differences is 0.050, while the average uncertainties (1σ) of the S^2 values are 0.020 and 0.017 in the isotropic and anisotropic cases, respectively. Thus, the S^2 values determined with the two diffusion models do not exhibit statistically significant differences.

Likewise, the chemical exchange rates, R_{ex} , determined on model-free analyses with the isotropic and axially symmetric diffusion models were compared (Fig. 21). The r.m.s. of the differences between the R_{ex} values with the two diffusion models is 0.75 s^{-1} , while the average uncertainties (1σ) of the R_{ex} values are 0.30 and 0.25 s^{-1} in the isotropic and anisotropic cases, respectively. This suggests that the R_{ex} values derived with both diffusion models do not exhibit statistically significant differences. Then, R_{ex} values determined on $J(0)$ analysis ($R_{\text{ex-j0}}$) were compared with R_{ex} values determined on model-free analysis ($R_{\text{ex-mf}}$) with the isotropic model (Figs. 18b and 21a). The average difference of $R_{\text{ex-j0}}$ values from $R_{\text{ex-mf}}$ values is $+0.75\text{ s}^{-1}$. This

positive difference arises mainly from the fact that the $R_{\text{ex-mf}}$ parameter was fixed at zero for dynamic models 1, 2 and 5. However, as the average uncertainty (1σ) of $R_{\text{ex-j0}}$ (0.84 s^{-1}) is comparable to the average difference of $+0.75\text{ s}^{-1}$, the chemical exchange rates determined on the present model-free analyses, $R_{\text{ex-mf}}$, are statistically reliable to the extent of their uncertainties (0.30 and 0.25 s^{-1} in the isotropic and anisotropic cases, respectively).

In the similar manner, the fast internal correlation times, τ_{ef} , with the two diffusion models were compared. The r.m.s. of the differences between the τ_{ef} values with the two diffusion models is 10.64 ps , which is comparable to the average uncertainty (11.37 ps) of the τ_{ef} values, showing that the τ_{ef} values determined with the two diffusion models are statistically similar (Fig. 22).

On model-free analysis with the isotropic diffusion model, model 5 was incorporated into the model selection, which includes the effective correlation times for slow internal motions, τ_{es} , in addition to the overall correlation time, τ_{m} . On the other hand, for analysis with the anisotropic rotational diffusion model, model 5 was not included in the model selection procedure because it has recently been shown that slow internal correlation times in an isotropic diffusion model often originate from the anisotropy of overall rotational diffusion (50). Out of the 27 spins assigned to model 5 in the isotropic model, 13 were fitted reasonably well to any of models 1 to 4 in the anisotropic model, only two could be fitted by the incorporation of the additional parameter, τ_{es} (model 5), and the remaining 12 spins could not be assigned to any of the dynamic models with the anisotropic model, probably because the orientations of the amide bonds for these 12 spins were not fixed rigidly (the average $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE value is 0.61 at 50.7 MHz). These results suggest that the slow internal motions of about the half of these 27 spins determined with the isotropic model may originate from the anisotropic overall motion, and the rest of the spins may actually have slow internal motions.

The results of analysis of the ^{15}N relaxation data showed that the rotational diffusion of the central domain of XPA has an anisotropic character. Thus, an anisotropic rotational diffusion model and an isotropic model were applied to model-free analyses of the ^{15}N relaxation data for the investigation of the internal motions of the backbone and the overall rotational diffusion of the domain. The result showed that the internal motion parameters obtained with the isotropic model and the anisotropic model are similar except for the parameters for the slow internal motions of nanosecond order. The large values of these slow internal motions determined on model-free analysis with the isotropic model may arise from the anisotropic overall motion.

Backbone dynamics and their biological implications

Uniformly high values of the order parameter, S^2 , are observed for the zinc-containing subdomain, except for N-terminal residues 98 to 103, suggesting that this region is among the most rigid ones in the protein molecule. This feature is probably due to the stabilization of the subdomain by the zinc coordination with associating extensive hydrogen bond networks, and the hydrophobic core formed by residues V103, F112 and M118 (22).

Small $\{^1\text{H}\}$ - ^{15}N NOE values (< 0.55 at 50.7 MHz), large T_2 values, and small S^2 values (< 0.6) are characteristic of a part of loop L1 (residues 158), a part of loop L2 (residues 168 to 179), the latter half of helix $\alpha 3$, and the following C-terminal region (residues 206 to 218), suggesting that these regions are not rigid but exhibit a large extent of mobility. These regions exhibited contributions from slow internal correlation times, τ_{es} , of an order of 1 ns on model-free analysis with the isotropic diffusion model, but they did not exhibit significant contributions from R_{ex} values. These observations suggest that these regions exhibit internal motions on a picosecond to nanosecond time scale. On the other hand, larger R_{ex} values ($> 0.2 \text{ s}^{-1}$ at 50.7 MHz) were observed for a part of loop L1 (residues 154 to 157), indicating the presence of conformational exchange on a microsecond to millisecond time scale. All these regions that exhibit picosecond to nanosecond motions or microsecond to millisecond motions are exposed to solvent, and are parts of the DNA binding surface, as suggested by the previous chemical shift perturbation experiment (22) (Fig. 16). The internal motions of these regions may be important for the interaction of the central domain of XPA with various kinds of damaged DNA by altering the conformation of the interaction surface to fit the structures of damaged DNAs. High rates of amide proton exchange are also observed in these regions (Fig. 7), suggesting that conformational exchange of large magnitude on a slower time scale is also possible.

The average T_1/T_2 ratio for the zinc-containing subdomain was almost the same as the average T_1/T_2 ratio for the C-terminal subdomain (Fig. 17c), showing that the two subdomains undergo rotational diffusion with similar correlation times (51). The hydrophobic core formed between residues Y116, L117, F121 and L123 in the zinc-containing subdomain, and residues L138, I165, L182 and L184 in the C-terminal subdomain, possibly restricts the independent motion of each subdomain (Fig. 12). The ^{15}N spins of residues 133 and 134 located in the linker sequence had large R_{ex} values ($> 3.5 \text{ s}^{-1}$ at 50.7 MHz), showing that the linker sequence may exhibit conformational exchange on a microsecond to millisecond time scale (Fig. 24).

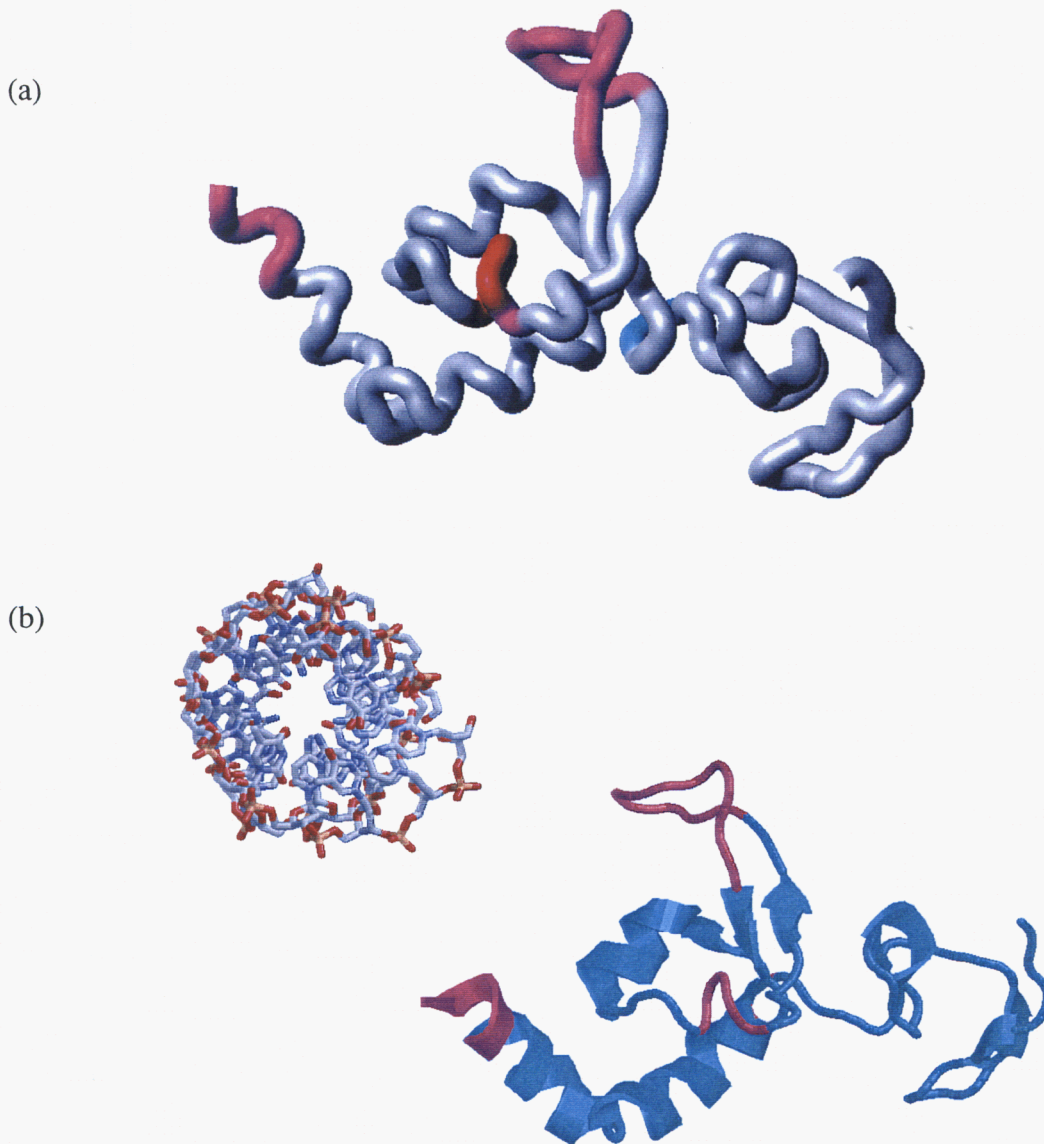


Fig. 24. Mapping of the flexible regions of XPA which are related to its function.

(a) Regions exhibiting internal motions on a picosecond to nanosecond time scale characterized by small $\{^1\text{H}\}$ - ^{15}N NOE values (< 0.55 at 50.7 MHz), large T_2 values, small S^2 values (< 0.6), slow internal correlation times of an order of 1 ns , and little contribution from R_{ex} values are drawn in magenta (residues 158, 168 to 179, and 206 to 218). A region exhibiting conformational exchange on a microsecond to millisecond time scale characterized by large R_{ex} values ($> 0.2 \text{ s}^{-1}$ at 50.7 MHz) is drawn in red (residues 154 to 157). These regions are in the interaction surface for DNA. Residues 133 and 134 with large R_{ex} values ($> 3.5 \text{ s}^{-1}$ at 50.7 MHz) are drawn in cyan. This region serves as a linker connecting the zinc-containing subdomain and the C-terminal subdomain, and exhibits conformational exchange on a microsecond to millisecond time scale. (b) The X-ray structure of a double-stranded DNA dodecamer containing a cisplatin adduct (78) is shown with a stick model. The flexible residues shown in (a) are also drawn in magenta on a ribbon model of XPA. The diameter of the DNA is almost the same as the width of the cleft on XPA.

CONCLUSION

NER is a major DNA repair pathway whose mechanism is highly conserved. However, the three-dimensional structures of none of the proteins involved in the early steps of NER were known, except for that of the single-stranded-DNA-binding domain of RPA70 (52). By means of NMR, I determined the three-dimensional structure of the central domain of human XPA, identified its interaction surfaces for DNA and RPA70, and analyzed its dynamic properties.

For the assignments of the chemical shifts of ^1H , ^{15}N , and ^{13}C nuclei, a series of heteronuclear and multi-dimensional NMR experiments was performed with the ^{15}N -, ^{15}N - ^{13}C -, or ^2H - ^{15}N - ^{13}C -labelled XPA₉₈₋₂₁₉. Almost all of the chemical shifts of the backbone and sidechain nuclei were assigned except for those of 10 residues.

For the structure calculations, the information on the interproton distances, dihedral angles and hydrogen bonds was obtained from the NOE spectra, scalar coupling constants, and amide hydrogen-deuterium exchange rates, respectively. The three-dimensional structures were calculated by the simulated annealing procedure with 1,336 interproton distance, 83 dihedral angle, and 47 hydrogen bond constraints. The calculated structures show that the central domain of XPA is composed of two subdomains, the zinc-containing subdomain with a compact globular structure and the C-terminal subdomain with a positively charged cleft.

For the identification of the DNA- and RPA70-binding surfaces of the central domain, chemical shift perturbation experiments were performed with the ^{15}N -labelled XPA₉₈₋₂₁₉. The results showed that the basic cleft in the C-terminal subdomain is involved in the DNA-binding, and the zinc-containing subdomain is involved in an RPA70-binding. This is one of the zinc-fingers which do not serve as DNA-binding surfaces, but interact directly with other proteins.

For the investigation of the dynamic properties of the central domain, the relaxation data (T_1 , T_2 , and heteronuclear NOE) of the amide ^{15}N spins were obtained. The analysis of the relaxation data showed that the domain has an anisotropic character of the rotational diffusion. Further model-free analyses with isotropic and anisotropic rotational diffusion models showed that the cleft in the C-terminal subdomain contains most of the flexible regions, and almost all of the zinc-containing subdomain is rigid. XPA recognizes structurally unrelated DNA damages such as (6-4) photoproducts and crosslinks caused by UV and chemicals like cisplatin, N-acetoxy-2-acetylaminofluorene, and osmium-tetroxide (8). The flexibility in the basic cleft may be important for the interactions of XPA with various kinds of damaged DNA.

The present structural data suggest that XPA interacts with RPA through the zinc-containing subdomain, and subsequently the complex binds to damaged DNA using the basic cleft in the C-terminal subdomain as an interface. By recruiting other factors to the damaged site, the XPA-RPA complex is supposed to serve as a platform for the formation of the repairsome for the early step of NER (10, 11, 13, 14, 17). Knowledge on the tertiary structure of the central domain of XPA, and its DNA- and RPA70-binding surfaces may facilitate understanding of the NER molecular assembly formed at damaged sites. It may also lead to an understanding of the serious genetic diseases caused by mutations in NER factors, in particular, xeroderma pigmentosum.

Coordinates and other NMR data

The coordinates have been deposited in the Brookhaven Protein Data Bank as *1xpa*. Other NMR data (chemical shifts, coupling constants, and distance and angle constraints) have been deposited in the BioMagResBank as *r1xpamr*.

REFERENCES

1. Aboussekhra, A. and Wood, R.D. (1994) Repair of UV-damaged DNA by mammalian cells and *Saccharomyces cerevisiae*. *Curr. Opin. Genet. Dev.* 4, 212-220.
2. Cleaver, J.E. and Kraemer, K.H. (1995) Xeroderma pigmentosum in *The Metabolic Basis of Inherited Disease, 7th Edn.* (C.R. Scriver *et al.*, eds.) pp. 4393-4419, McGraw-Hill, New York.
3. Hoeijmakers, J.H. and Bootsma, D. (1990) Molecular genetics of eukaryotic DNA excision repair. *Cancer Cells Mon. Rev.* 2, 311-320.
4. Tanaka, K. *et al.* (1990) Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. *Nature* 348, 73-76.
5. Robins, P., Jones, C.J., Biggerstaff, M., Lindahl, T., and Wood, R.D. (1991) Complementation of DNA repair in xeroderma pigmentosum group A cell extracts by a protein with affinity for damaged DNA. *EMBO J.* 10, 3913-3921.
6. Jones, C.J. and Wood, R.D. (1993) Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA. *Biochemistry* 32, 12096-12104.
7. Asahina, H. *et al.* (1994) The XPA protein is a zinc metalloprotein with an ability to recognize various kinds of DNA damage. *Mutat. Res.* 315, 229-237.
8. Kuraoka, I., Morita, E.H., Saijo, M., Matsuda, T., Morikawa, K., Shirakawa, M., and Tanaka, K. (1996) Identification of a damaged-DNA binding domain of the XPA protein. *Mutat. Res.* 362, 87-95.
9. He, Z., Henricksen, L.A., Wold, M.S., and Ingles, C.J. (1995) RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* 374, 566-569.
10. Li, L., Lu, X., Peterson, C.A., and Legerski, R.J. (1995) An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Mol.*

Cell. Biol. 15, 5396-5402.

11. Saijo, M., Kuraoka, I., Masutani, C., Hanaoka, F., and Tanaka, K. (1996) Sequential binding of DNA repair proteins RPA and ERCC1 to XPA *in vitro*. *Nucleic Acids Res.* 24, 4719-4724.
12. Park, C.-H., Mu, D., Reardon, J.T., and Sancar, A. (1995) The general transcription-repair factor TFIIH is recruited to the excision repair complex by the XPA protein independent of the TFIIIE transcription factor. *J. Biol. Chem.* 270, 4896-4902.
13. Nocentini, S., Coin, F., Saijo, M., Tanaka, K., and Egly, J.-M. (1997) DNA damage recognition by XPA protein promotes efficient recruitment of transcription factor II H. *J. Biol. Chem.* 272, 22991-22994.
14. Li, L., Elledge, S.J., Peterson, C.A., Bales, E.S., and Legerski, R.J. (1994) Specific association between the human DNA repair proteins XPA and ERCC1. *Proc. Natl. Acad. Sci. USA* 91, 5012-5016.
15. Li, L., Peterson, C.A., Lu, X., and Legerski, R.J. (1995) Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Mol. Cell. Biol.* 15, 1993-1998.
16. Nagai, A. *et al.* (1995) Enhancement of damage-specific DNA binding of XPA by interaction with the ERCC1 DNA repair protein. *Biochem. Biophys. Res. Commun.* 211, 960-966.
17. Mu, D., Hsu, D.S., and Sancar, A. (1996) Reaction mechanism of human DNA repair excision nuclease. *J. Biol. Chem.* 271, 8285-8294.
18. Wakasugi, M. and Sancar, A. (1998) Assembly, subunit composition, and footprint of human DNA repair excision nuclease. *Proc. Natl. Acad. Sci. USA* 95, 6669-6674.
19. Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H. (1998) Xeroderma pigmentosum group C protein

- complex is the initiator of global genome nucleotide excision repair. *Mol. Cell* 2, 223-232.
20. Morita, E.H., Ohkubo, T., Kuraoka, I., Shirakawa, M., Tanaka, K., and Morikawa, K. (1996) Implications of the zinc-finger motif found in the DNA-binding domain of the human XPA protein. *Genes Cells* 1, 437-442.
21. Miyamoto, I., Miura, N., Niwa, H., Miyazaki, J., and Tanaka, K. (1992) Mutational analysis of the structure and function of the xeroderma pigmentosum group A complementing protein. Identification of essential domains for nuclear localization and DNA excision repair. *J. Biol. Chem.* 267, 12182-12187.
22. Ikegami, T., Kuraoka, I., Saijo, M., Kodo, N., Kyogoku, Y., Morikawa, K., Tanaka, K., and Shirakawa, M. (1998) Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nature Struct. Biol.* 5, 701-706.
23. Cavanagh, J., Fairbrother, W.J., Palmer III, A.G., and Skelton, N.J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego.
25. Hu, W. and Zuiderweg, E.R.P. (1996) Stereospecific assignments of Val and Leu methyl groups in a selectively ^{13}C -labeled 18 kDa polypeptide using 3D CT-(H)CCH-COSY and 2D $^1\text{J}_{\text{C}}\text{C}$ edited heteronuclear correlation experiments. *J. Magn. Reson. B* 113, 70-75.
26. Hu, J.-S., Grzesiek, S., and Bax, A. (1997) Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $\text{J}_{\text{C}'\text{C}\gamma}$ and $\text{J}_{\text{NC}\gamma}$ couplings. *J. Am. Chem. Soc.* 119, 1803-1804.
27. Brünger, A.T. (1993) *X-PLOR 3.1: a system for X-ray crystallography and NMR*. Yale University Press, New Haven, Connecticut.
28. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14, 51-55.
29. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J.M.

- (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477-486.
30. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33, 5984-6003.
31. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277-293.
32. Garrett, D.S., Powers, R., Gronenborn, A.M., and Clore, G.M. (1991) A common sense approach to peak picking in two-, three, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* 95, 214-220.
33. Mandel, A.M., Akke, M, and Palmer, III A.G. (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246, 144-163.
34. Bevington, P.R. and Robinson, D.K. (1992) *Data reduction and error analysis for the physical sciences 2nd eds.* pp. 38-52, 161-164, McGraw-Hill, USA.
35. Peng, J.W. and Wagner, G. (1995) Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry* 34, 16733-16752.
36. Wong, K.-B., Fersht, A.R., and Freund, S.M.V. (1997) NMR ¹⁵N relaxation and structural studies reveal slow conformational exchange in barstar C40/82A. *J. Mol. Biol.* 268, 494-511.
37. Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104, 4546-4559.

38. Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104, 4559-4570.
39. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112, 4989-4991.
40. Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. (1996) Anisotropic rotational diffusion of perdeuterated HIV protease from ¹⁵N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR* 8, 273-284.
41. Woessner, D.E. (1962) Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys.*, 37, 647-654.
42. Wishart, D.S. and Sykes, B. (1994) Chemical shifts as a tool for structure determination in *Methods in Enzymology* (James, T.L. and Oppenheimer, N.J., eds.) Vol. 239, pp. 363-392, Academic Press, San Diego.
43. Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol.* 233, 123-138.
44. Omichinski, J.G. *et al.* (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* 261, 438-446.
45. Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352, 497-505.
46. Summers, M.F. (1996) Zinc fingers in *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D.M. and Harris, R.K., eds.) Vol. 8, pp. 5063-5071, John Wiley and Sons Ltd., West Sussex, England.

47. Satokata, I., Iwai, K., Matsuda, T., Okada, Y., and Tanaka, K. (1993) Genomic characterization of the human DNA excision repair-controlling gene XPAC. *Gene* 136, 345-348.
48. Go, M. (1985) Protein structures and split genes. *Adv. Biophys.* 19, 91-131.
49. Nakashima, T. *et al.* (1994) Structure of the human CCG1 gene : relationship between the exons/introns and functional domain/modules of the protein. *Gene* 141, 193-200.
50. Schurr, J.M., Babcock, H.P., and Fujimoto, B.S. (1994) A test of the model-free formulas. Effects of anisotropic rotational diffusion and dimerization. *J. Magn. Reson.* B105, 211-224.
51. Zhou, H., McEvoy, M.M., Lowry, D.F., Swanson, R.V., Simon, M.I., and Dahlquist, F.W. (1996) Phosphotransfer and CheY-binding domains of the histidine autokinase CheA are joined by a flexible linker. *Biochemistry* 35, 433-443.
52. Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., and Frappier, L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385, 176-181.
53. Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
54. Merritt, E.A. and Bacon, D.J. (1997) Raster3D photorealistic molecular graphics. *Methods Enzymol.* 277, 505-524.
55. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* 11, 281-296.
56. Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., and Frappier, L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385, 176-181.

57. Pfuetzner, R.A., Bochkarev, A., Frappier, L., and Edwards, A.M. (1997) Replication protein A : Characterization and crystallization of the DNA binding domain. *J. Biol. Chem.* 272, 430-434.
58. Moggs, J.G., Yarema, K.J., Essigmann, J.M., and Wood, R.D. (1996) Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1,3-intrastrand d(GpTpG)-cisplatin adduct. *J. Biol. Chem.* 271, 7177-7186.
59. Jones, C.J., Cleaver, J.E., and Wood, R.D. (1992) Repair of damaged DNA by extracts from a xeroderma pigmentosum complementation group A revertant and expression of a protein absent in its parental cell line. *Nucleic Acids Res.* 20, 991-995.
60. McDowell, M.L., Nguyen, T., and Cleaver, J.E. (1993) A single-site mutation in the XPAC gene alters photoproduct recognition. *Mutagenesis* 8, 155-161.
61. Guzder, S.N., Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1995) Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIIH. *J. Biol. Chem.* 270, 12973-12976.
62. Aboussekhra, A. *et al.* (1995) Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* 80, 859-868.
63. Pfuetzner, R.A., Bochkarev, A., Frappier, L., and Edwards, A.M. (1997) Replication protein A : Characterization and crystallization of the DNA binding domain. *J. Biol. Chem.* 272, 430-434.
64. Lee, K.M., Androphy, E.J., and Baleja, J.D. (1995) A novel method for selective isotope labeling of bacterially expressed proteins. *J. Biomol. NMR* 5, 93-96.
65. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) Gradient-tailored water suppression for ^1H - ^{15}N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson.* A102, 241-245.
66. Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. (1993) Two-dimensional NMR experiments for correlating ^{13}C β and ^1H δ / ϵ chemical shifts of aromatic residues in ^{13}C -labeled proteins via

scalar couplings. *J. Am. Chem. Soc.* 115, 11054-11055.

67. Sattler, M., Schmidt, P., Schleucher, J., Schedletsky, O., Glaser, S.J., and Griesinger, C. (1995) Novel pulse sequences with sensitivity enhancement for in-phase coherence transfer employing pulsed field gradients. *J. Magn. Reson.* B108, 235-242.

68. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* 85, 393-399.

69. Kay, L.E., Xu, G.-Y., Singer, A.U., Muhandiram, D.R., and Forman-Kay, J.D. (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^1H and ^{13}C correlations in H_2O samples of proteins. *J. Magn. Reson.* B101, 333-337.

70. Muhandiram, D.R. and Kay, L.E. (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson.* B103, 203-216.

71. Shaka, A.J., Keeler, J., Frenkiel, T., and Freeman, R. (1983) An improved sequence for broadband decoupling: Waltz-16. *J. Magn. Reson.* 52, 335-338.

72. Shaka, A.J., Lee, C.J., and Pines, A. (1988) Iterative schemes for bilinear operators; application to spin decoupling. *J. Magn. Reson.* 77, 274-293.

73. Ikegami, T., Sato, S., Wälchli, M., Kyogoku, Y., and Shirakawa, M. (1997) An efficient HN(CA)NH pulse scheme for triple-resonance 4D correlation of sequential amide protons and nitrogens-15 in deuterated proteins. *J. Magn. Reson.* 124, 214-217.

74. Rajagopal, P., Waygood, E.B., Reizer, J., Saier, Jr.M.H., and Klevit, R.E. (1997) Demonstration of protein-protein interaction specificity by NMR chemical shift mapping. *Protein Science* 6, 2624-2627.

75. Thompson, L.H. (1998) Nucleotide excision repair in *DNA damage and repair, Volume II: DNA repair in higher eukaryotes* (Nickoloff, J.A. and Hoekstra, M.F. eds.) pp. 335-393, Human

Press, New Jersey.

76. Sarkar, S.K. (1996) *NMR spectroscopy and its application to biomedical research*, Elsevier Science B.V., Amsterdam.

77. Reid, D.G. (1997) Protein NMR techniques in *Methods in Molecular Biology* (Walker, J.M. eds.) Vol. 60, Human Press, New Jersey.

78. Takahara, P.M., Rosenzweig, A.C., Frederick, C.A., and Lippard, S.J. (1995) Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. *Nature* 377, 649-652.

79. Kobayashi, T., Takeuchi, S., Saijo, M., Nakatsu, Y., Morioka, H., Otsuka, E., Wakasugi, M., Nikaido, O., and Tanaka, K. (1998) Mutational analysis of a function of xeroderma pigmentosum group A (XPA) protein in strand-specific DNA repair. *Nucleic Acids Res.* 26, 4662-4668.

List of Publications

1. Ikegami, T., Kuraoka, I., Saijo, M., Kodo, N., Kyogoku, Y., Morikawa, K., Tanaka, K., and Shirakawa, M. (1998) Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nature Struct. Biol.* 5, 701-706.
2. Ikegami, T., Sato, S., Wälchli, M., Kyogoku, Y., and Shirakawa, M. (1997) An efficient HN(CA)NH pulse scheme for triple-resonance 4D correlation of sequential amide protons and nitrogens-15 in deuterated proteins. *J. Magn. Reson.* 124, 214-217.
3. Ikegami, T., Kuraoka, I., Saijo, M., Kodo, N., Kyogoku, Y., Morikawa, K., Tanaka, K., and Shirakawa, M. (1998) Resonance assignments, solution structure, and backbone dynamics of the DNA- and RPA-binding domain of human repair factor XPA. *J. Biochem.* (in press).

ACKNOWLEDGEMENTS

The present studies have been performed under the direction of Dr. Yoshimasa Kyogoku (Institute for Protein Research, Osaka University), and Dr. Masahiro Shirakawa (Graduate School of Biological Sciences, Nara Institute of Science and Technology). I would like to express my gratitude to them for their guidance and discussions throughout my studies.

I wish to thank Dr. Isao Kuraoka, Dr. Masafumi Saijo, Dr. Naohiko Kodo, and Dr. Kiyoji Tanaka (Institute for Molecular and Cellular Biology, Osaka University) for the gifts of the genes of the central domain of human XPA and its C153S mutant and purified RPA₁₈₁₋₄₂₂ protein, and for the precious discussions.

I wish to thank Dr. Markus Wälchli (Bruker Japan) for many setups and practical advice for the NMR measurements.

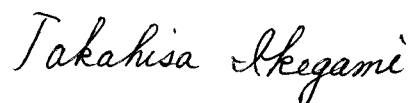
I wish to thank Ms. Mie Maeda, and Dr. Kosuke Morikawa (Biomolecular Engineering Research Institute) for the helpful discussions.

Thanks to Dr. Toshio Hakoshima and all the members of his group (Graduate School of Biological Sciences, Nara Institute of Science and Technology), I could carry out my experiments comfortably. I would like to appreciate their kindness.

During my studies, I received various helps from many persons. I am grateful to Dr. Young Ho Jeon (Biotech Research Institute, LG Chemical Ltd.), Dr. Masato Shimizu (Biomolecular Engineering Research Institute) and Dr. Hayato E. Morita (Department of Chemistry, Ehime University) for their advice and instructions about the preparations of the samples.

Finally, I particularly thank my wife, Keiko, for her patience, support, and understanding during my research period.

Takahisa Ikegami



November 18, 1998.

謹呈