Intramolecular interactions in pancreatic ribonucleases

ELENA YU. KOLBANOVSKAYA,¹ BANGALORE K. SATHYANARAYANA,²
ALEXANDER WLODAWER,² AND MARAT YA. KARPEISKY¹

¹ V.A. Engelhardt Institute of Molecular Biology, Academy of Sciences of Russia, Vavilova Str. 32, Moscow, Russia

² Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center,
ABL-Basic Research Program, Frederick, Maryland 21702

(RECEIVED December 27, 1991; REVISED MANUSCRIPT RECEIVED March 27, 1992)

Abstract

A detailed analysis of the composition and properties of hydrophobic nuclei and microclusters in pancreatic ribonuclease A (RNase A) has been carried out. Distance calculations for all noncovalently bonded atoms revealed that the average number of nonpolar contacts between a side chain of an amino acid and its neighbors is substantially larger if it involves hydrophobic residues rather than nonhydrophobic ones. However, the difference decreased when the number of contacts per nonpolar group and/or atom were calculated. Three main nuclei and five microclusters were identified, and their quantitative parameters were calculated. These nuclei include hydrophobic residues with a substantial number of nonpolar contacts with the environment (Phe 8, Phe 120, Phe 46, Tyr 25, Tyr 97, Ile 107, Leu 35, Ile 81, Val 54, Val 108, Met 29, Met 30). Hydrophobic nuclei of RNase A differ in shape and in composition, in the number of intranuclear contacts and of associated residues, as well as in their internal mobility. All eight cysteine residues are involved in nonpolar interactions with amino acid residues of hydrophobic nuclei. Active site amino acid residues of RNase A form a noncovalent contact network comprised of themselves, as well as of many conserved residues from hydrophobic nuclei. Sequence alignment with some other members of the RNase A family of proteins shows remarkable similarity in positions and in conservation of the main nonpolar residues, comprising cores of two (out of three) hydrophobic nuclei. A correlation was shown to exist between the average density of contacts for side-chain atoms and the number of amino acids to be found in the appropriate positions in the sequences of related mammalian ribonucleases. However, there are certain amino acid positions in the third, smaller nucleus, which are highly variable within the family. Taking into account that this nucleus is composed of residues belonging to different elements of the secondary structure, it is likely that the mutual orientation of these elements can be somehow different for these proteins.

Keywords: hydrophobic nuclei; nonpolar contacts; ribonuclease A

The family of enzymes that includes bovine pancreatic ribonuclease (EC 3.1.27.5, RNase A) has been studied very extensively and provides one of the best model systems to investigate structure–function relationship in proteins, including folding and the structural basis of biological activity. Enzymes belonging to that family have been objects of detailed physicochemical and biological studies during the last several decades (for reviews see Richards & Wyckoff, 1971; Blackburn & Moore, 1982; Wlodawer, 1984; Harris et al., 1987; Benner & Allemann, 1989; De Llorens et al., 1989). The active site, as well as a detailed description of the secondary structure, including the hydrogen-bonding network system, have been described af-

ter analyses of high resolution structures of RNase A (Fig. 1) and of its complexes with nucleotides in the crystalline state (Wlodawer & Sjölin, 1983; Borkakoti et al., 1984; Wlodawer, 1984; Borah et al., 1985; McPherson et al., 1986; Harris et al., 1987; Wlodawer et al., 1988) as well as in solution (Karpeisky & Yakovlev, 1981; Udgaonkar & Baldwin, 1988; Jimenez et al., 1990). However, less information has been obtained from these structural data with respect to assignment, composition, and characterization of nonpolar regions of RNase A, since quantitative approaches for delineation of nonpolar regions of a protein were formulated only recently (Heringa & Argos, 1991; Karpeisky & Ilyin, 1992).

We have undertaken a study of the distribution and parameters of noncovalent contacts in the RNase A and related proteins in order to rationalize the similarity in catalytic properties of these enzymes with sequences that

Reprint requests to: Marat Ya. Karpeisky, V.A. Engelhardt Institute of Molecular Biology, Academy of Sciences of Russia, Vavilova Str. 32, Moscow, Russia.

Fig. 1. Ribbon diagram of the three-dimensional structure of RNase A.

are homologous but not identical. In the meantime, amino acid sequences of more than 40 pyrimidine-specific ribonucleases became known, and certain conclusions on the sequence-secondary structure relationship were drawn on the basis of multiple alignment (Beintema &

van der Laan, 1986; Beintema, 1987; Beintema et al., 1988; Schuller et al., 1988). Amino acid sequence homology for proteins under consideration was found to be more than 50%, and consequently they should have similar folding (Chothia & Lesk, 1986; Sander & Schneider, 1991). The validity of this assumption will be discussed later, based on the results of our computations. To visualize noncovalent contacts in the RNase A molecule we took advantage of the "contact map" approach based on distance calculation for all noncovalently bonded atoms in a protein (Rossmann & Liljas, 1974; Nemethy & Scheraga, 1979; Plochocka et al., 1988; Sander & Godzic, 1989; Karpeisky et al., 1990; Karpeisky & Ilyin, 1992).

Results and discussion

Secondary structure

The contact map of ribonuclease A, calculated from the atomic coordinates of phosphate-free RNase A (7RSA; Wlodawer et al., 1988) at a cutoff distance of 5.2 Å, is shown in Figure 2A. Regular elements of the structure, such as α -helices (residues 3–13, 24–34, 50–60), are seen as a thickening along the main diagonal of the map. β -Strands are represented by contacts arranged in a linear fashion normal to the main diagonal (residues 41–48, 61–63, 71–75, 79–87, 94–110, 118–123), whereas reverse β -turns (residues 22–25, 36–39, 65–68, 75–78, 87–90,

Fig. 2. Contact maps of the noncovalent contacts in phosphate-free RNase A at a cutoff distance of 5.2 Å; residues that were found in multiple conformations are considered in their A form (Wlodawer et al., 1988; PDB file 7RSA). Contact pairs (i,j) are indicated by marks. **A:** Map of all noncovalent contacts. Numbers of contacts between residues (i,j) are indicated by: cross, 2–5; circle, 6–15; filled circle, 16–25; filled square, >25. Open square, S–S bridges. **B:** Map of nonpolar contacts. Filled square, contacts between residues (i,j) that include at least two contacts between nonpolar groups and/or atoms of side chains; open square, contacts between residues that do not involve nonpolar atoms and/or groups.

91–94, 112–115) are detected on the map as crosslike contact areas located closely to the main diagonal. Thus, the delineation obtained of the elements of regular secondary structure in RNase A agreed well with previously published interpretations based on the analysis of torsion angles and hydrogen bonds in the structure (Borkakoti et al., 1984; Harris et al., 1987; Wlodawer et al., 1988).

Contacts between nonpolar groups

An interaction between nonpolar groups of protein side chains is described quantitatively by its interaction energy. An approximate measure of the extent of interaction, however, can be given by specifying the number of pairs of atoms or of residues that are in contact (Chou et al., 1985). To that end, we carried out an analysis of distances between the side-chain nonpolar groups (CH, CH₂, CH₃, and S) of all amino acid residues in the RNase A molecule, based on crystallographically derived atomic coordinates (Fig. 2B). One can assume that such residues as Phe, Ile, Leu, and Met, having more than 20 nonpolar contacts per residue side chain, should be located in central parts of protein nonpolar regions, and consequently they have to be preserved in three-dimensional structure of other members of the family. However, the difference looks much less remarkable when the number of contacts per carbon atom was calculated (Table 1). Many of the nonpolar contacts involve residues well separated in amino acid sequence. In addition, the nonpolar environment of the residue is not correlated di-

Fig. 3. Schematic diagram of the nonpolar contact network of RNase A at a cutoff distance of 4.5 Å. Residues *i* and *j*, connected by straight lines, have at least two pairs of nonpolar side-chain groups and/or atoms in contact. Marked circles, either invariant amino acids in sequences of RNase A family proteins, or a given position can be occupied only by a closely related amino acid.

Table 1. Nonpolar contacts of side-chain carbon atoms of RNase A, at a cutoff distance of 5.2 Å

Amino acid composition per molecule of RNase A	Total number of contacts for side-chain carbon atoms	Average number of contacts per residue	Average number of contacts per side-chain carbon atom
Ala 12	50	4.2	4.2
Cys 8	121	15.1	7.6
Asp 5	22	4.4	4.4
Glu 5	40	8.0	4.0
Phe 3	135	45.0	6.4
His 4	55	13.8	4.6
Ile 3	83	27.7	6.9
Lys 10	51	5.1	1.3
Leu 2	47	23.5	5.9
Met 4	146	36.5	9.1
Asn 10	22	2.2	2.2
Pro 4	59	14.8	4.9
Gln 7	69	9.9	5.0
Arg 4	36	9.0	3.0
Ser 15	37	2.5	2.5
Thr 10	75	7.5	3.8
Tyr 6	163	27.2	4.5
Val 9	156	17.3	5.8

rectly with the nature of its side chain (e.g., see V108 and V43 in Fig. 2B). These data are in line with the known fact that globular proteins have high packing densities, which are readily attainable among clusters of the naturally occurring hydrophobic amino acid residues (Behe et al., 1991). In accordance with the proposed definitions as described in the Methods section, the nonpolar environment for every RNase A residue has been identified (data not shown). An analysis of the composition and structure of the nonpolar contact network of each residue in a family resulted in construction of a scheme of nonpolar contacts among residues. These data were used to reveal nonpolar regions and to identify three main nuclei, as well as five microclusters (Fig. 3). These nuclei include classic hydrophobic residues with "populated" families, namely Phe 8, Phe 120, Phe 46, Tyr 25, Tyr 97, Ile 107, Leu 35, Ile 81, Val 54, Val 108, Met 29, and Met 30 (Table 2). We further assigned Gln 11 to the first nucleus and Leu 35 to the third. These hydrophobic nuclei of RNase A (Table 3; Figs. 3, 4) are different in shape (number of "triangles"), in composition, in the number of intranuclear contacts, and in the number of associated residues.

These nuclei were found to include not merely the classical hydrophobic amino acids, but also dicarboxylic acids (Glu 49, Glu 86, Asp 121), their amides (Gln 74, Asn 62, Gln 60, Gln 55, Asn 44), as well as basic amino acids (Arg 33, Lys 41, Lys 104, His 12). All eight cysteine residues are involved in nonpolar interactions with amino acid residues of hydrophobic nuclei. Some quantitative characteristics of these nuclei are listed in Table 3.

Table 2. *Hydrophobic nuclei (γ) and microclusters (mc) of RNase A calculated from file 7RSA^a*

$\gamma 1$	F8; Q11*; H12; M13; T45; V47; E49; L51; V54; Q55; V57; C58; Y73; M79; L81; A102; K104; I106; V108; C110; Y115; P117; F120 (E9; N62; S75; V116; S123)
$\gamma 2$	K61; V63; C65; C72; Q74; Y76; I107 (Q60; Q69; D121; A122; V124)
$\gamma 3$	A20; Y25; C26; M29; M30; R35*; L35*; K41; F46; T82; C84; Y97 (R10; P42; N44; H48; E86; T99)
mc1	A4; E111; V118
mc2	T36; C40; Y92; P93; C95
mc3	V43; R85
mc4	T78; H105
mc5	A109; H119

^a RNase A, phosphate free, resolution 1.26 Å; d-min = 3.0 Å; d-max = 4.5 Å; minimal number of contacts between residues (K) = 2; minimal number of nonpolar neighbors = 2. Residues associated with a nucleus due to contacts with one nonpolar neighbor are italicized. Residues that comprise a bridge between the first and the third nuclei are marked by *

A commonly used approach to assigning a residue to a hydrophobic cluster is based on calculations of either the accessibility of a residue to water molecules (Richards, 1977; Plochocka et al., 1988; Lesser & Rose, 1990) or molecular area of a residue (Finney, 1978). That information is necessary, but not sufficient, to study the topology, structure, and properties of nonpolar regions of a protein. Nonpolar side chains were found occasionally to be located on the surface of a protein globule, being involved in numerous contacts with buried residues simultaneously. If our approach to assigning a residue to hydrophobic regions of a protein is correct, then these residues will be identified by other methods as well. To that end the average values of static accessibility (E) and of molecular area (Am) for amino acid residues assigned by us to nuclei were calculated. The average <E> for the components of the nuclei $\gamma 1$ and $\gamma 3$ were found to be 15%, and for $\gamma 2$, 27%. (The corresponding <Am> values are 14.2 Å² for the nucleus $\gamma 1$, 15.5 Å² for $\gamma 3$, and 32.3 Å² for $\gamma 2$.) The average <E> for all other residues (located neither in nonpolar nuclei nor in the active site area) is 50% (mean <Am> for these residues is 25.5 Å²). The anomalous low average value of <E> (26%) was found even for polar and charged residues situated in the nuclei of RNase A.

The temperature factor values were used to characterize the atomic mobility in hydrophobic nuclei (Table 3). According to the averaged values, the nuclei should be considered as relatively rigid parts of a protein globule. The temperature factors for the atoms that belong to the first and third nuclei have rather low values, close to those of atoms in α -helices (mean = 9.8 Å²) or in β -

sheets (mean = 9.2 Å²). The nucleus $\gamma 2$, located near the surface of a protein globule (Fig. 4B), is less compact and its constituents are more mobile (Table 3). The most compact part of the main hydrophobic nucleus $\gamma 1$ (Fig. 4A) comprises the residues of the first α -helix, interacting with residues of the third α -helix, and their mutual orientation is stabilized due to the interactions with Tyr 73, Met 79, Ile 106, and Val 108 of the principal β -strands (residues 71–92 and 94–110), as well as with Phe 120 and Val 47 of two other short β -strands (residues 41–48 and 118–123). The second hydrophobic nucleus, $\gamma 2$, in which the central residue is Ile 107, is relatively small, less compact, and consists of the residues found in the central part of the polypeptide chain of RNase A. The two nuclei are connected by a bridge resulting from the nonpolar contacts of Val 57 ($\gamma 1$) with Gln 60 ($\gamma 2$). The hydrophobic residues of the second α -helix are involved in interactions with hydrophobic residues of the same β -strands, namely Tyr 97 and Phe 46, to form the third nucleus ($\gamma 3$). The nonpolar contact bridge between the first and the third nucleus was found to involve the N-terminal α -helix residue, Gln 11 (Fig. 4C).

In addition, the average density of contacts was calculated both for side chains of residues comprising nonpolar nuclei as well as for all the other residues (see Methods; Heringa & Argos, 1991; Karpeisky & Ilyin, 1992). The data thus obtained (Table 4) show that the average density of contacts per residue in nonpolar nuclei considerably exceeds that of other residues. In other words, the residues of nuclei form numerous contacts with the environment, and that determines their functional and structural importance.

Table 3. *Quantitative characteristics of hydrophobic nuclei of RNase A*

	$\gamma 1$	$\gamma 2$	$\gamma 3$
N (c.p.) ^a	98	16	56
N (t) ^b	81	23	45
N (c/at) ^c	2.42	1.39	2.49
N (c/res.) ^d	8.52	4.57	9.33
N (n) ^e	2.6	2.0	2.5
T/nuc. ^f	0.13	0.14	0.25
(A.r.)/R ^g	0.23	0.71	0.50
 ^h	12.0	16.5	11.6

^a Number of contacting atom pairs (sulfur and united atoms).

^b Total number of nonpolar contacting atoms.

^c Average number of contacts per contacting atom.

^d Average number of contacts per residue.

^e Average number of neighbors per residue.

^f Average number of triangles per residue.

^g Average number of associated residues per residue.

^h Average value of temperature factor for side-chain hydrocarbon atoms (sulfur atoms included).

Fig. 4. Mutual orientation of γ nuclei and secondary structure elements in RNase A. Atoms in each nonpolar group of the residue included in any nuclei are presented as van der Waals spheres. **A:** Nucleus $\gamma 1$. **B:** Nucleus $\gamma 2$. **C:** Nucleus $\gamma 3$.

Group of residues ^a	$\Sigma_n(C_k)^b$	D^b	$\Sigma_n(C_{knp})^b$	D_{np}^b
$\gamma 1$ ($n = 23$)	21.40	0.93	6.47	0.28
$\gamma 2$ ($n = 7$)	5.83	0.83	1.50	0.21
$\gamma 3$ ($n = 12$)	12.20	1.02	3.77	0.31
In ($n = 23$)	20.29	0.88	2.32	0.10
Vn ($n = 56$)	26.03	0.47	1.83	0.03

^b $\sum(C_k)$, D , $\sum(C_{knn})$, D_{nn} , for definitions see Methods.

The comparison of variability of amino acid residues, comprising the nonpolar nuclei in RNase A, with that of other residues (except for active site residues, which are identical in all proteins of the family) reveals the nonpolar nuclei to be the most conserved regions in the molecules of pancreatic mammalian RNases. The average variability for amino acid residues in the nuclei, taking into account generally accepted conservative substitutions (Risler et al., 1988), is equal to 4% for the nucleus $\gamma 1$, 16% for $\gamma 2$, and 5% for $\gamma 3$. This value was estimated to be equal to 19% for the other residues. These data suggest that the composition and contact network of the nucleus 2, as well as those of the hydrophobic microclusters of the RNase A-like proteins, should tolerate the substi-

One can assume that the static accessibility (E) of a residue belonging to the hydrophobic nucleus should correlate to some extent with the number of amino acids to be found in the appropriate positions in the sequences of related ribonucleases. Calculated values of accessibility of amino acids (Lee & Richards, 1971) increase, passing from invariant for the whole RNase A family residues ($N = 1$, where N corresponds to the number of different residues found at a given position in aligned sequences of the RNase A family of proteins), through residues with one or two substitutions ($N = 2$ or 3), to those with multiple substitutions ($N = 4$ or more) (Fig. 5). When N becomes relatively large, the value of E approaches a constant level (about 55–65%). As mentioned above, average static accessibility of the side-chain atoms of the residues in hydrophobic nuclei was found to be equal to 15% for $\gamma 1$ and $\gamma 3$ and to 27% for $\gamma 2$. However, there are certain positions in the polypeptide chain of RNase A occupied by different amino acids in other RNases (Table 5). As a rule these positions are localized in exposed turns and loops (mean $E = 59\%$), and corresponding residues form a few nonconservative contacts with the environment (less than six pairs per residue). It is reasonable to assume that these nonspecific contacts should be dispensable, and they can be established due to any type of side-chain interactions. The invariant residues involved in reverse β -turns, namely Thr 36, Asn 67, Ser 75, Pro 93, and Gly 112, are supposed to be of importance for maintaining the polypeptide chain conformation of the enzyme. Thus, it can be concluded that the differences in sequence for the proteins in the family affect neither the composition nor the nonpolar network structure of

```

?  ~  ~  ~  ?  ~  ~  !  ~  ~  ~  ?  ~  ?  ?  ?
m  *  ***  *  **  **  *  m  m  *  m  ***  *  *  *  *  *
KETA AAKFERQHMDSS TSAASSSN YCNM M KSRNLTKDRCKPVTN FVHESLADVQAVCSQKNVA
1      2      4      6
      0      0      0
~  ~  ~  ?  ?  ?  ?  ~  ~  !  ~  ~  ~  ?  !  ?  ~  ~  ~  ?  ?  ~  ~  ~
*  ***  *  m  *  *  m  m  m  *  *  *  m  *  m  *  *  m  *
CKNGQTNCYQSYSTM L TDCRETGSSKY PNCAYKTTQANKHII VACEGNPYVPVHFDASV
      8      1      1
      0      0      2
      0      0      0

```

^a *, Amino acid residues composing hydrophobic nuclei of RNase A; m, amino acid residues composing a hydrophobic microcluster; ^, invariant amino acids of RNase A family proteins; ~, position can be occupied by a closely related hydrophobic amino acid; !, position can be occupied by a closely related charged amino acid; ?, variable position. Comment marks ^, ~, !, and ? are included only for those amino acid residues that are part of either nuclei or microclusters, as defined in the text.

Fig. 5. Correlation between static accessibilities of residues in RNase A and the number of substitutions found in aligned sequences of the RNase A family of proteins. N, number of different amino acids found in the particular location; E(%), mean value of static accessibilities of residues with a given N.

RNase A. Consequently, a postulate of identical folding, from which this work was started, can be considered as being quite plausible.

Active site region

Various definitions of the enzyme active site have been discussed in the past, and the most appropriate one refers to a region of the protein globule where the substrate molecule is bound, and enzyme-substrate complexes are

Fig. 6. Correlation between average densities of contacts of side-chain atoms of RNase A residues and the number of substitutions found in aligned sequences of the RNase A family of proteins. N, number of different amino acids found in the particular location; D, average density of contacts of side-chain atoms of residues with a given N.

formed and transformed (Karpeisky, 1976). It comprises molecules of immobilized water in addition to amino acid residues. Active site amino acid residues of RNase A form a noncovalent contact network comprised of themselves, as well as of residues from nonpolar nuclei. Thus, 16 contact pairs involved only the active site residues, whereas 71 contact pairs involved both the active site residues and residues from their environment. In addition to hydrogen bonding, there are a certain number of nonpolar contacts involving active site residues and those residues that are part of hydrophobic nuclei and microclusters (Fig. 3; Table 2). The pyrimidine-specific substrate binding cleft of the RNase A active site (Gln 11, Phe 120, Thr 45, Val 47) includes amino acid residues that belong mostly to the nucleus γ 1. However, catalytic residues (His 12, His 119, Lys 41, Asp 121) belong to different nuclei and microclusters (Fig. 3). One can assume that nonpolar contacts

Table 6. Alignment of the sequences of RNase A and of related proteins^a

	BFL	EDN	RNase (h/b)	AGN (h/b)	P-30
-1		K	-	-	
-2		P	-	-A	
-3		P	-	-Q	
1	-	Q	K	DD	-
2	-	F	E	ND	-
α -Helix H1					
3	E	T	ST	SY	E
4	N	?	RA	R	D
5	W	A	A	Y	W
6	A	Q	KA	TI	L
7	T	W	K	H	T
8	F	F	Fn!	F	F
9	Q	E	QE	L	Q
10	Q	T	R	T	K
11	K	Q	Q	Q	K
12	H	H	Hn!	H	H
13	I	I	Mn*	Y	I
End					
14	-	N	D	D	-
15	I	M	S	A	T
16	N	T	DS	K	N
17	T	-	ST	P	T
18	P	-	S	QK	R
19	I	-	PA	G	D
20	I	-	SAn?	R	V
21	N	-	S	DN	D
22	-	-	S	DD	-
23	-	S	S	RE	-
α -Helix H2					
24	-	Q	TN	-	-
25	-	Q	Yn!	Y	-
26	C	C	Cn!	C	C
27	N	T	N	EF	D
28	T	N	Q	SH	N
29	I	A	Mn!	IM	I
30	M	M	Mn!	MM	M
31	D	Q	K	RK	S
32	N	V	RS	RN	T
33	N	I	Rn*	RR	N
34	I	N	N	GR	L

(continued)

Table 6. Continued

	BFL	EDN	RNase (h/b)	AGN (h/b)	P-30		BFL	EDN	RNase (h/b)	AGN (h/b)	P-30
End						<i>β-Strand (continued)</i>					
35	Y	N	MLn*	L	F	85	T	N	Rm*	K	-
36	I	Y	Tm!	T	-	86	R	L	LE	LH	-
37	V	Q	K	SR	-	87	T	T	T	HK	-
38	G	R	GD	P	-	88	S	T	NG	G	N
39	G	R	R	-	H	89	I	P	GS	G	V
39a	Q					90	T	S	S	S	T
40	C	C	Cm!	C	C	91	P	P	RK	PS	S
<i>β-Strand</i>						92	R	Q	Ym*	WR	R
41	K	K	Kn!	K	K	End					
42	R	N	P	D	D	93	P	N	Pm!	P	P
43	V	Q	Vm!	IR	K	<i>β-Strand</i>					
44	N	N	N!	N	N	94	-	I	N	P	-
45	T	T	Tn!	T	T	94a	N				
46	F	F	Fn!	F	F	94b	S				
47	I	L	Vn*	I	I	95	C	C	Cm!	C	C
48	I	L	H	H!	Y	96	P	R	A	QR	K
End						97	Y	Y	Yn!	Y	Y
49	S	T	En!	G	S	98	S	A	RK	RG	K
<i>α-Helix H3</i>						99	S	Q	T	A	L
50	S	T	PS	N	R	100	R	T	ST	T	K
51	A	F	Ln!	K	P	101	T	P	PQ	AE	K
52	T	A	VA	RN	E	102	E	A	KAn?	QD	S
53	T	N	D	SD	P	103	T	N	EN	FS	T
54	V	V	Vn*	I	V	104	N	M	RKn*	R	N
55	K	V	Qn?	K	K	105	Y	F	Hm?	V	K
56	A	N	NA	A	A	106	I	Y	In*	VI	F
57	I	V	Vn*	I	I	107	C	I	In*	V	C
58	C	C	Cn!	C	C	108	V	V	Vn*	V	V
59	T	G	FS	KE	K	109	K	A	Am!	AG	T
60	-	N	Q	ND	-	110	C	C	Cn!	C	C
End						End					
<i>β-Strand</i>						111	E	D	Em*	E	E
61	G	P	EKn?	KR	G	112	N	N	G	N	N
62	V	N	KN	N	I	113	Q	R	SN	G	Q
63	I	M	Vn*	G	I	113a	-	D	-	-	-
64	-	T	TA	NQ	A	113b	-	Q	-	-	-
65	N~	C	Cn!	P~	S	113c	-	R	-	-	-
End						113d	-	R	-	-	-
66	M	P	K	HY	K	113e	-	D	-	-	-
67	N	S	N	R	N	113f	-	P	-	-	-
68	V	N	G	EG	V	113g	-	P	-	-	-
69	L	K	Q	-	L	113h	-	Q	-	-	-
70	S	T	GT	-	T	113i	-	Y	-	-	-
70a	R					114	-	P	P	-	-
70b	K					115	-	V	Yn!	-	-
<i>β-Strand</i>						116	Y	V	Vn*	L	A
71	-	N	N	ND	-	117	P	P	Pn!	P	P
72	-~	C	Cn!	L~	-~	<i>β-Strand</i>					
73	-	H	Yn!	R	-	118	V	V	Vm!	V	V
74	-	H	KQ	I	-	119	H	H	Hm!	H	V
75	T	S	S	S	T	120	F	L	Fn!	LF	F
76	-	G	NYn?	K	-	121	A	D	D	D	V
77	T	S	S	S	S	122	G	R	A	GE	G
78	R	Q	STm?	SE	E	123	I	I	S	S	V
79	F	V	Mn*	F	F	End					
80	Q	P	HS	Q	Y	124	G	I	V	IF	
81	L	L	In*	VI	L		R	-	E~	FI	S
82	N	I	Tn!	T	S		C	-	D~	RT	C
83	T	H	D	TI	D		V	-	S~	RP	-
84	C	C	Cn!	C	C		-	-	--	PR	-
					(continued)		-	-	--	-H	-

^a Numbering corresponds to that of bovine pancreatic ribonuclease A. In pancreatic RNase A, S-S bonds link residues 26-84, 40-95, 58-110, and 65-72. His 12, Lys 41, and His 119 are at the active site. BFL, bullfrog lectin (Kamiya et al., 1990); EDN, eosinophil-derived neurotoxin (Gleich et al., 1986); RNase, pancreatic ribonuclease A: human-h, bovine-b (Beintema, 1987); AGN, angiogenin: h-human, b-bovine (Bond & Strydom, 1989); P-30, anti-tumor protein (Ardelt et al., 1991). n, Residue belongs to a nonpolar nucleus; m, residue belongs to nonpolar microcluster; !, invariant residue in the RNase A family; *, semiconservative residue in RNase A family; ?, variable residue in the RNase A family.

between substrate binding and catalytic amino acids provide necessary conditions for stabilization of the enzyme-substrate complex in a catalytically competent conformation. Many of the functionally important noncovalent contacts involving active site residues were found to be preserved in the whole family of RNase A-related proteins (Table 5). The results indicate that difference in sequence should not affect composition and three-dimensional structure of the active site. To understand whether or not the number of contacts of a residue in the RNase A molecule determines its variability in the family, one has to compare the average density of contacts for side-chain atoms with the number of amino acids to be found in the appropriate positions in the sequences of related ribonucleases. The results of such calculations (Fig. 6) clearly indicate that the residues with numerous side-chain contacts in the three-dimensional structure of RNase A tend to be the ones most conserved in the family.

RNase A-related proteins with specific biological properties

Sequence alignment of some proteins with important biological properties that are able to catalyze hydrolysis of RNA (Table 6) shows remarkable similarity in positions and conservation of active site residues as well as of main nonpolar residues of RNase A. These nonpolar residues have the most populated families (Fig. 2B) and they contribute significantly into structural organization of the hydrophobic nuclei (Fig. 3). The average variability for amino acid residues for proteins listed in Table 6 is equal to 23% for the nucleus $\gamma 1$, 20% for $\gamma 3$, and 57% for $\gamma 2$. The other residues (except those of the active site) have the average variability of 52%. A large number of invariant and/or conservative residues belong either to the first or to the third hydrophobic nucleus (Phe 8, Val 47, Val 54, Val 57, and Pro 117 of $\gamma 1$; and Met 30, Phe 46, and Tyr 97 of $\gamma 3$), suggesting that they should play a crucial role in spatial organization of the proteins under consideration. The second nucleus of RNase A ($\gamma 2$) was concluded to represent the most variable part of the protein nonpolar region, being capable of tolerating different substitutions. However, Val 63 and Ile 107 were semiconservative in the family of proteins discussed here. Taking into account substantial number of nonpolar contacts between themselves and with their environment, one has to assign them an important role in maintaining structural integrity of the protein globule. Three out of four S-S bridges found in the RNase A molecule (58–110, 26–84, 40–95) can exist in other proteins as well. The cysteine residues of the fourth one (65–72) are involved in formation of an exposed loop and are not conserved in RNase A-related proteins, with an exception of eosinophil-derived neurotoxin (EDN). However, there are many positions, especially in the nucleus $\gamma 2$, occupied by amino

acid residues that could not support nonpolar contacts. This indicates that there are certain differences in the three-dimensional structure of RNase A and those of RNase A-like proteins.

The analysis thus performed suggests that the RNase A-related proteins should have similar folding, but mutual orientation of regular structural elements does not have to be identical. That means that differences in sequence do not affect global folding, but conformation and composition of exposed loops should not be the same for different proteins with RNase A-like catalytic properties, and one can speculate that biological activity of these proteins is associated with position, sequence, and conformation of these loops.

Methods

The structures of RNase used in this study were taken from the Protein Data Bank (Bernstein et al., 1977). They include the model refined jointly with 2.0-Å X-ray and neutron data (5RSA; Wlodawer & Sjölin, 1983), an independently refined RNase model at 1.45 Å resolution (6RSA; Borkakoti et al., 1984), and the structure of phosphate-free RNase at 1.26 Å resolution (7RSA; Wlodawer et al., 1988).

A contact pair was defined to be a pair of any amino acid residues, provided that they had at least two atom pairs separated by the distance equal to or less than 5.2 Å, as previously discussed (Karpeisky & Ilyin, 1992). Other distances were also used for specific calculations. Coordinates of side-chain hydrocarbon atoms were used as coordinates of "united atoms"—CH, CH₂, CH₃. The number of contacts was computed on the basis of atomic van der Waals radii according to a method described earlier (Karpeisky & Ilyin, 1992).

The analysis has been carried out using the following definitions:

1. *The hydrophobic family of a residue.* A number of amino acid residues selected on the basis that each of them has at least two nonpolar groups (either hydrocarbon group and/or sulfur atoms) separated from analogous groups of the residue and/or atoms under consideration by a distance equal to or less than 5.2 Å (the criterion of a nonpolar neighbor). It should be kept in mind that the number of hydrophobic families determined by this method is far greater than the actual number of clusters because, in terms of the criterion given above, different families include one and the same amino acids.

2. *Library.* A list of hydrophobic families for every residue of a protein under consideration, arranged according to the sequence of the protein.

3. *Nonpolar region of a protein globule.* A part of a protein globule comprising amino acid residues selected out of the library such that each of them has at least one nonpolar neighbor from the library.

4. *Hydrophobic cluster*. The part of a nonpolar region in a protein in which every residue has two or more nonpolar neighbors at the cutoff distance of 5.2 Å.

5. *Hydrophobic nucleus*. The most compact part of a nonpolar region of the protein globule comprised of amino acid residues having not less than two nonpolar neighbors at a cutoff distance of 4.5 Å, provided that at least three of them contact each other to form an interacting triangle.

6. *Hydrophobic microcluster*. A group of amino acid residues, in which every residue has at least one hydrophobic neighbor at a cutoff distance of 4.5 Å. The residues involved do not form interacting triangles. At a cutoff distance of 4.5 Å, certain residues in nonpolar regions of a protein were found to have only one nonpolar neighbor. These residues either form microclusters or can be associated with a nucleus.

A scheme of nonpolar contacts among residues was built up on the basis of hydrophobic library data. A visual inspection of the scheme allows one to make definite conclusions regarding composition of the nonpolar regions of the protein molecule. On the basis of the scheme of nonpolar regions and in accordance with definitions 5 and 6 given above, the composition and pattern of interactions among residues of hydrophobic nuclei and microclusters were defined.

The average density of contacts of side-chain atoms (D) was calculated by the formula:

$$D = \frac{\sum_k (C_k)}{n},$$

where C_k is the normalized sum of contacts between side-chain atoms of k th residue (k corresponds to the number of residue according to the RNase A amino acid sequence) and all atoms of the other residues, and n is the number of residues in the group of residues under consideration (for a hydrophobic nucleus).

$$C_k = \frac{\sum_j (A_{kj})}{NORM(TYPE(R_i))} \quad j = 1, 2, \dots, 124; \quad j \neq k,$$

where $NORM(TYPE(R_i))$ is the average maximum value of side-chain atom contacts for a particular residue type that are within 4.5 Å (Heringa & Argos, 1991), and A_{kj} is the number of surrounding main-chain and side-chain atoms of residue j that are within 4.5 Å of side-chain atoms of residue k . Analogously, the average density of nonpolar contacts (D_{np}) and the corresponding C_{knp} was calculated (Heringa & Argos, 1991; Karpeisky & Ilyin, 1992).

The composition of nuclei and of microclusters was not affected by small differences in atomic coordinates of

RNase A deposited in the Protein Data Bank (1RN3, 5RSA, and 7RSA; data not shown). The differences affected slightly the number of contacts between pairs and their distribution pattern in nuclei only, but did not change any of the final conclusions.

Acknowledgments

We acknowledge support from the Institute of Molecular Biology, Academy of Sciences of Russia, obtained while A.W. was a visiting scientist at that institution. M.Y.K. has been a recipient of a visiting fellowship from the National Cancer Institute during his stay in Frederick. These mutual visits made this work possible. This research was sponsored in part by the National Cancer Institute, DHHS, under contract N01-CO-74101 with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

- Ardelt, W., Mikulski, M., & Shogen K. (1991). Amino acid sequence of an anti-tumor protein from *Rana pipiens* oocytes and early embryos. Homology to pancreatic ribonucleases. *J. Biol. Chem.* 266, 245–251.
- Behe, M.J., Lattman, E.E., & Rose, G.D. (1991). The protein-folding problem: The native fold determines packing, but does packing determine the native fold? *Proc. Natl. Acad. Sci. USA* 88, 4195–4199.
- Beintema, J.J. (1987). Structure, properties and molecular evolution of pancreatic type ribonucleases. *Life Chem. Rep.* 4, 333–389.
- Beintema, J.J., Schuller, C., Irie, M., & Carsana, A. (1988). Molecular evolution of the ribonuclease superfamily. *Progr. Biophys. Mol. Biol.* 51, 165–192.
- Beintema, J.J. & van der Laan, J.M. (1986). Comparison of the structure of turtle pancreatic ribonuclease with those of mammalian ribonucleases. *FEBS Lett.* 194, 338–342.
- Benner, S.A. & Allemann R.K. (1989). The return of pancreatic ribonucleases. *Trends Biochem. Sci.* 14, 396–397.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Mayer, E.F., Jr., Bryce, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 122, 535–542.
- Blackburn, P. & Moore, S. (1982). Pancreatic ribonuclease. In *The Enzymes*, Vol. XV, Part B (Boyer, P.D., Ed.), pp. 317–433. Academic Press, New York.
- Bond, M.D. & Strydom, D.Y. (1989). Amino acid sequence of bovine angiogenin. *Biochemistry* 28, 6110–6113.
- Borah, B., Chi-Wan Chen, Egan, W., Miller, M., Wlodawer, A., & Cohen, J.S. (1985). Nuclear magnetic resonance and neutron diffraction studies of the complex of ribonuclease A with uridine vanadate, a transition-state analogue. *Biochemistry* 24, 2058–2067.
- Borkakoti, N., Moss, D.S., Stanford, M.J., & Palmer, R.A. (1984). The refined structure of ribonuclease-A at 1.45 Å resolution. *J. Crystallogr. Spectrosc. Res.* 14, 467–494.
- Chothia, C. & Lesk, A.M. (1986). The relation between the divergence of the sequence and structure in proteins. *EMBO J.* 5, 823–826.
- Chou, K.C., Nemethy, G., Rumsey, S., Tuttle, R.W., & Scheraga, H.A. (1985). Interactions between an α -helix and β -sheets. Energetics of α/β packing in proteins. *J. Mol. Biol.* 186, 591–609.
- De Llorens, R., Arus, C., Pares, X., & Cuchillo, C.M. (1989). Chemical and computer graphics studies on the topography of the ribonuclease A active site cleft. A model of the enzyme–pentanucleotide substrate complex. *Protein Eng.* 2, 417–429.
- Finney, J.L. (1978). Volume occupation, environment, and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* 119, 415–441.

- Gleich, G.J., Loegering, D.A., Bell, M.P., Checkel, S.J., Ackerman, S.J., & McKean D.J. (1986). Biochemical and functional similarities between human eosinophil-derived neurotoxin and eosinophil cationic protein: Homology with ribonuclease. *Proc. Natl. Acad. Sci. USA* 83, 3146–3150.
- Harris, G.W., Borkakoti, N., Moss, D.S., Palmer, R.A., & Howlin, B. (1987). Ribonuclease A. Analysis of the hydrogen bond geometry and spatial accessibility at the active site. *Biochim. Biophys. Acta* 912, 348–356.
- Heringa, J. & Argos, P. (1991). Side chain clusters in protein structures and their role in protein folding. *J. Mol. Biol.* 220, 151–171.
- Jimenez, M.A., Rico, M., Herranz, J., Santoro, J., & Nieto, J.L. (1990). Solution structure of the isolated ribonuclease C-terminal 112–124-fragment. *Biochim. Biophys. Acta* 1038, 322–329.
- Kamiya, Y., Oyama, F., Oyama, R., Sakakibara, F., Nitta, K., Kawauchi, H., Takayanagi, Y., & Tinani, K. (1990). Amino acid sequence of a lectin from Japanese frog (*Rana japonica*) eggs. *J. Biochem.* 108, 139–143.
- Karpeisky, M.Ya. (1976). Enzymic active site: Stereochemistry and dynamics. *Mol. Biol. (USSR)* 10, 1197–1210.
- Karpeisky, M.Ya. & Ilyin, V.A. (1992). Analysis of nonpolar regions in proteins. *J. Mol. Biol.* 5, 629–638.
- Karpeisky, M.Ya., Kolbanovskaya, E.Yu., & Borisov, V.V. (1990). Intramolecular interactions in pancreatic ribonucleases. *DAN USSR* 315, 495–499.
- Karpeisky, M.Ya. & Yakovlev, G.I. (1981). Topochemical principles of substrate specificity of nucleases. *Sov. Sci. Rev. D2*, 145–257.
- Lee, B. & Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.
- Lesser, G.J. & Rose, G.D. (1990). Hydrophobicity of amino acid subgroups in proteins. *Proteins Struct. Funct. Genet.* 8, 6–13.
- McPherson, A., Brayer, G.D., & Morrison, R.D. (1986). Crystal structure of RNase A complexed with d(pA₄). *J. Mol. Biol.* 189, 305–327.
- Nemethy, G. & Scheraga, H.A. (1979). A possible folding pathway of bovine pancreatic RNase. *Proc. Natl. Acad. Sci. USA* 76, 6050–6054.
- Plochocka, D., Zielenkiewicz, P., & Rabczenko, A. (1988). Hydrophobic microdomains as structural invariant regions in proteins. *Protein Eng.* 2, 115–119.
- Richards, F.M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6, 151–176.
- Richards, F.M. & Wyckoff, H.W. (1971). Bovine pancreatic ribonuclease. *Enzymes IV*, 647–806.
- Risler, J.L., Delorme, M.O., Delacroix, H., & Henaut, A. (1988). Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J. Mol. Biol.* 204, 1019–1029.
- Rossmann, M.G. & Liljas, A. (1974) Recognition of structural domains in globular proteins. *J. Mol. Biol.* 85, 177–181.
- Sander, C. & Godzik, A. (1989). Conservation of residue interactions in a family of Ca-binding proteins. *Protein Eng.* 2, 589–596.
- Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins Struct. Funct. Genet.* 9, 56–68.
- Schuller, C., Neuteboom, B., Wubbels, G.H., Beintema, J.J., & Nevo, E. (1988). The amino acid sequence of pancreatic ribonuclease from the mole rat *Spalax ehrenbergi*, chromosomal species 2n = 60. In *Molecular Evolution of Pancreatic Ribonuclease from Rodents*, pp. 79–94. Drukkerij van Denderen, B.V., Groningen.
- Udgaonkar, J.B. & Baldwin, R.L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335, 694–699.
- Wlodawer, A. (1984). Structure of bovine pancreatic ribonuclease by X-ray and neutron diffraction. In *Biological Macromolecules and Assemblies, Nucleic Acids and Interactive Proteins* (Jurnak, F.A. & McPherson, A., Eds.), pp. 394–440. Wiley, New York.
- Wlodawer, A. & Sjölin, L. (1983). Structure of ribonuclease A: Results of joint neutron and X-ray refinement at 2.0 Å resolution. *Biochemistry* 22, 2720–2728.
- Wlodawer, A., Svensson, L.A., Sjölin, L., & Gilliland, G.L. (1988). Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry* 27, 2705–2717.