ADVANCES IN ENZYMOLOGY

AND RELATED AREAS OF MOLECULAR BIOLOGY

Volume 34

ADVANCES IN ENZYMOLOGY

AND RELATED AREAS OF MOLECULAR BIOLOGY

Edited by F. F. NORD

FORDHAM UNIVERSITY, BRONX, NEW YORK

VOLUME 34

1971
INTERSCIENCE PUBLISHERS
a division of John Wiley & Sons, Inc.
New York · London · Sydney · Toronto

Library of Congress Catalog Card Number: 41-9213

ISBN 0 471 64963 5

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

FUNCTION OF AMINO ACID SIDE CHAINS*

By JOHANN MATHEJA† and EGON T. DEGENS, Woods Hole, Massachusetts

CONTENTS

* Contribution No. 2471 from the Woods Hole Oceanographic Institution. This work was supported by a grant from the National Aeronautics and Space Administration (NASA-2-014-001).

† Present address: Department of Physical Chemistry, Kernforschungsanlage Jülich 1, West Germany.

Part I Serine

I. Introduction

The primary structure of a number of polypeptides is known, and substitution in homologous proteins is used for comparative biochemistry. It has been shown that the sequential order of amino acids in enzymes is related to phylogeny. This type of study has biochemical implications, since one can examine which amino acid residues in a peptide chain remain invariant towards evolutionary changes.

It is well known that enzymes function as (*1*) catalyst and (*2*) control devices by means of molecular recognition and thus have a duplex nature. When molecules resemble the substrate molecules with respect to shape, they are bound to the active centers, thereby retarding the enzyme activity. This represents an isosteric association because the competitive inhibitors resemble the substrate molecule. In allosteric regulations, on the other hand, molecules with no structural or chemical resemblance to the enzyme substrate become bonded. The specific bonding introduces a conformational change which may activate or deactivate the enzymes.

Since the geometric shape of an enzyme is related to the chain of amino acids present, the type of amino acid side chain sequences exposed along the enzyme molecule influences the regulation mechanism for enzymic activities. In the present work we discuss the structural aspects of serine as they relate to enzymic activity; this amino acid is found in the active center of many enzymes. For a recent review on the chemistry of enzymes see reference 1.

II. Codon Assignments

Formula: $C_3H_7NO_3$; $^+H_3N—\underset{\displaystyle CH_2OH}{\underset{|}{CH}}—COO^-$

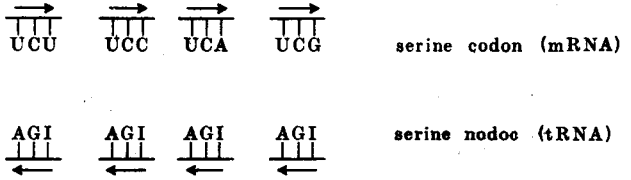
Molecular weight: 105.1 (serine residue, MW = 87.1 = 144×10^{-24} g)

Codon triplets: UCU; UCC; UCA; UCG; AGU; AGC.

Serine has six codon assignments of which the UC combination is used most frequently. Due to the high number of possible triplets—six—serine is better protected against random mutation than amino

acids with fewer triplet assignments. Arginine and leucine are the only other amino acids with a choice of six triplets.

The primary structure of serine transfer-RNA has been determined by Zachau and co-workers (2), who isolated this molecule from baker's yeast. The anticodon—frequently also termed nodoc—has the most likely base sequence IGÀ:



The presence of inosine (I) opposite to all four triplets in the third position provides a biochemical advantage, simply because this specific complementary base pairing of inosine reduces the number of tRNA's. The location of the tRNA's for the AGU and AGC codons is uncertain since it is unlikely that there would be two anticodon sites at the same transfer molecule. Two tRNA's (tRNA 1 and tRNA 2) have been isolated that differ in 3 nucleotides in a total of 85 bases. Aside from the four common bases (U, A, C, G), 10 different rare nucleosides have been identified in the serine tRNA's.

III. Function of Serine Side Chain

A. OCCURRENCE

Serine is quite common in peptides and proteins. It is of special significance that serine is abundant in structural proteins such as silk or conchiolin, as well as in functional proteins (enzymes). A ratio of Gly:Ala:Ser = 3:2:1 is described for the β-configuration of silk fibroin, and serine supplants alanine in the layer structure in concentrations up to 15 mole % (3). In the shell matrix protein of molluscs (conchiolin), which is operational in mineralization processes, the level for serine may reach concentrations as high as 20 mole % and is commonly around 10 mole %; this represents the highest percentage for serine observed in proteins (4). The values for serine in some

TABLE

Peptide Sequence next to Enzyme

Enzymes	
Proteolytic and esterase enzymes	
Chymotrypsinogen A	Bovine pancreas
Chymotrypsinogen B	Bovine pancreas
Chymotrypsinogen C	Porcine pancreas
Elastase	Porcine pancreas
Trypsinogen	Bovine pancreas
Thrombin	Bovine serum
Subtilisin (BPN′)	*Bacillus subtilis*
Subtilisin (Carlsberg)	*Bacillus subtilis*
Subtilisin (NOVO)	*Bacillus subtilis*
Aspergillopeptidase B	*Aspergillus oryzae*
Aspergillopeptidase F_1	*Aspergillus flavus*
Caseinase	French bean
Acetylcholinesterase	*torpedo* (electric organ)
Ali esterase	Equine liver
Carboxyl esterase	Equine liver
Pseudocholinesterase	Equine serum
Peptidase	Baker's yeast
Peptidase (Phaseolain)	French bean (*Phaseolus vulgaris*)
Phosphorylating enzymes	
Alkaline phosphatase	Bovine intestine
Alkaline phosphatase	*Serratia marcescens*
Alkaline phosphatase	*Escherichia coli*
Phosphoglucomutase	*Escherichia coli*
Phosphoglucomutase	Flounder
Phosphoglucomutase	Rabbit muscle
Phosphoglucomutase	Rabbit muscle
Phosphorylase	Rabbit muscle

representative enzymes and proteins are listed below in mole %:

Papain papaya	Trypsinogen bovine	Cytochrome	Ribonuclease	Hemoglobin
6%	15%	0–8%	15%	4–9%

The relationship between primary structure and the properties of proteins has been studied in detail and much has been learned. However, it is still impossible to predict the properties of a peptide chain in relation to its primary structure; the single problem of deciding which part of a peptide sequence in the protein chain may become important for enzyme activity still eludes us. Solely on empirical grounds it has been established that the serine is part of the active center in hydrolases,

I

Active Sites (Serine) in Hydrolases

Sequence	Ref.
Gly Val Ser Ser Cys Met Gly Asp SER Gly Gly Pro Leu Val Cys Lys Lys Asn	28
Gly Val Ser Ser Cys Met Gly Asp SER Gly Gly Pro Leu Val Cys Gln Lys Asn	29
Gln Thr Cys Glu Gly Asp SER Gly (Gly Pro) Leu Val	30
Gly Val Arg Ser Gly Cys Gln Gly Asp SER Gly Gly Pro Leu His Cys Leu Val Asn	13, 31
Gly Gly Lys Asp Ser Cys Gln Gly Asp SER Gly Gly Pro Val Val Cys Ser Gly Lys	32
Gly Asp SER Gly (Glu, Ala)	33
Lys Tyr Gly Ala Tyr Asn Gly Thr SER Met Ala Ser Pro His Val Ala Gly Ala	34
Thr Tyr Ala Thr Leu Asn Gly Thr SER Met Ala Ser Pro His Val Ala Gly Ala	34
Asn Gly Thr SER Met Ala Ser Pro His	34
Thr SER Met Ala	35
Gly Thr SER Met Ala	35
Thr SER Met Ala	36
Glu SER Ala	37
Gly Glu SER Ala Gly Gly (Glu, Ser)	38
Phe Gly Glu SER Ala Gly Gly	33
Phe Gly Glu SER Ala Gly (Ala, Ala, Ser)	39
Glu SER Val	36
Glu SER Val	36
Asp SER Ala	40
Thr Asp SER Ala	41
Thr Gly Lys Pro Asp Tyr Val Thr Asp SER Ala Ala Ser Ala	42
Thr Ala SER His Asn	44
Thr Ala SER His Asp Pro Gly Gly Pro Asp Asp Gly Phe	44
Thr Ala SER His Asx (Pro Gly Gly Pro Asx Asx Gly) Phe	45
Glu (Leu, Asp) Gly Val Thr Ala SER His Asp Gly Glu Ser Ala Gly Leu Asp Leu	46
Lys Gln Ile SER Val Arg	41

and characteristic amino acid sequences are observed next to the active sites (Table I).

The primary structure of known proteins frequently reveal typical serine successions:

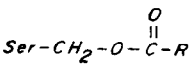
Ser-Ser; Ser-Ser-Ser; Ser-X-Ser; Ser-X-Y-Ser

We do not know why serine occurs in these combinations.

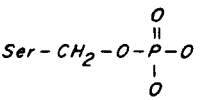
B. ABILITY OF SERINE TO FORM HYDROGEN BONDS

Serine has a terminal OH group which is separated by only one CH_2 unit from its backbone peptide chain. Such a small distance of approximately 2.5 Å between the active side group and the peptide chain

occurs in only one other case, that of cystine, $-CH_2-SH$. Chemical reactions such as esterification involving serine, or interactions of the

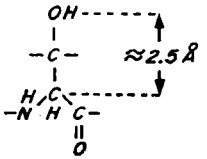


acylation



esterification

serine OH group with other molecules capable of generating hydrogen bonds, have a great affect on the conformation of the peptide chain. This results principally from the small distance between the active side group of serine and the peptide backbone:

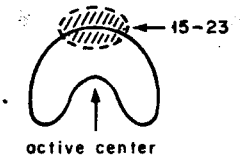


serine

A change of the molecular environment around the OH group of the serine promotes conformational changes and thus alters the secondary and tertiary structure of the peptide chain; part of the enzymic dynamics of serine is a consequence of this feature.

Diffraction analysis of the tertiary structure of ribonuclease (bovine) reveals (5) that the serine residues occur in those parts of the protein chain that form the surface of the molecule. The sequence:

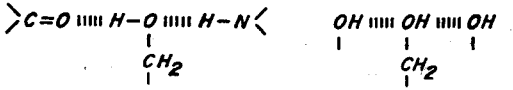
-Ser-Ser-Thr-Ser-Ala-Ala-Ser-Ser-Ser-
15 23



shape of ribonuclease (bovine)

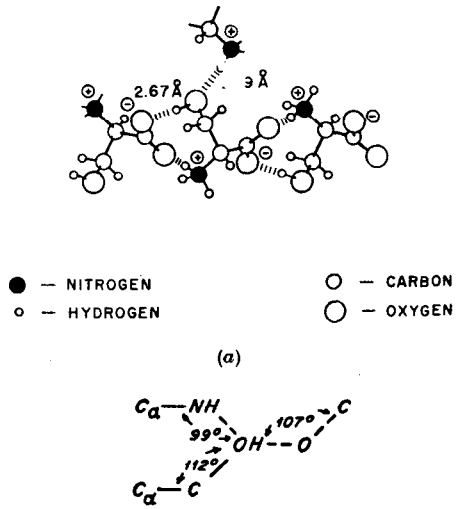
forms a loop at the surface of the ribonuclease which is ready to incorporate "structural" water. Judging by this characteristic, it appears that the loop influences the configuration of peptide chains. It is even conceivable that it plays a role in conformational change during enzyme activity. The folding pattern of the protein chain of ribonuclease has the shape of a horseshoe; the hydrated serine loop determines the position of the two ends of the "horseshoe" surrounding the active enzyme center.

The main function of serine, however, lies in its ability to form hydrogen bridges which increase the degree of crosslinkage among protein chains and raise their hydrophobic properties. The oxygen of the hydroxyl group can form two hydrogen bridges by supplying one H atom and attracting another one:



This process can go so far that serine supports the dispersion bonding forces between the $-CH_3$ and $-CH_2-$ residues by bringing these groups in close contact. Silk fibroins, which are remarkably resistant to chemical influences, are a good example of the ability of serine to form crosslinks.

The characteristics of hydrogen bridges established by the side chains of serine are evident from its crystal structure (6). A segment of the serine structure is shown in Figure 1. The hydroxyl group



Bonding angles of H-bridges and angles of the serine ($C\alpha-CH_2-OH$) residue chain.

Figure 1 Hydrogen bonds generated by the OH group of serine: (*a*) Segment of the molecular packing of a serine crystal, the dotted lines represent H bonds. (*b*) Bonding angle between the H bonds and the side chain.

participates in two hydrogen bridges; the placement of the OH group between the charged amino and carboxyl residues is of special interest.

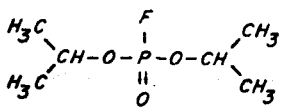
Serine is a main factor in the formation of tertiary structures whenever it is present in high concentration (>10 mole %). H bond formation by serine yields 10–15 kcal/mole if the hydroxyl groups at the hydrogen bond linkage are placed in an electrostatic field developed by ions, as is the case in the structure of serine crystals:

$$\overset{\oplus}{NH}\cdots\overset{\ominus}{OH}\overset{\oplus}{\cdots}O^{\ominus}$$
serine

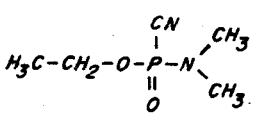
C. THE ACTIVE SERINE SITES IN ENZYMES

Serine is an active side chain residue of two kinds of enzymes: (*a*) proteolytic enzymes including esterases and (*b*) enzymes responsible for the transfer of phosphate groups.

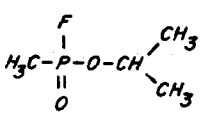
The presence of unstable side groups in peptides and proteins, which easily undergo chemical reaction, was recognized some years ago. The heating of protein will readily release the OH group of serine. In the history of the biosciences it is rather ironical that the recognition of serine as an active part of hydrolases was a product of wartime research. During World War II, extremely toxic nerve gases (organophosphoester) were developed and considered for use in chemical–biological warfare. In England, the gas diisopropylfluorophosphate (DFP) was developed; the deadly dose for humans is 500 mg. In Germany, the nerve gas Tabun with a toxicity 10 times that of DFP was manufactured; so was Sarin, which is even more deadly. Initially, organophosphoesters



di-isopropylfluorophosphate (DFP)



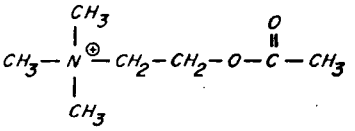
tabun



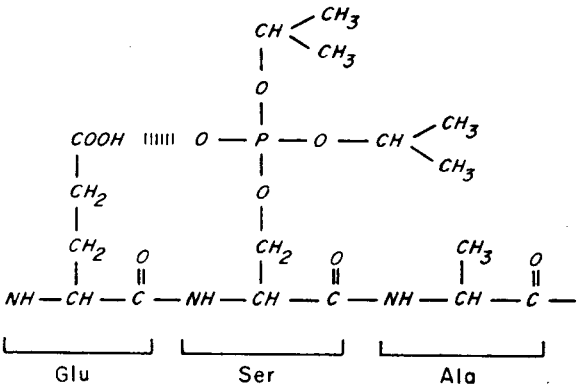
sarin

were used for pesticides and insecticides. Biochemical research has shown the extraordinary neuro-toxic action of organophosphoesters in connection with the enzyme cholinesterase activity. This enzyme hydrolyzes the chemical messenger acetylcholine that is involved in nerve impulse transmission. Consequently, the action of nerve gas is due to an irreversible phosphorylation of the active centers so that the enzyme activity becomes inhibited. Molecular biological studies have

revealed that serine represents that part of the cholinesterase which attaches the phosphate group:



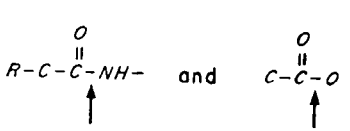
acetylcholine



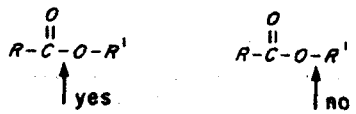
DFP phosphorylated serine in cholinesterase

Additional information was obtained by the study of other enzymes. Diisopropylfluorophosphate (DFP) reacts with esterases and proteases, and this nerve gas is now a biochemical reagent widely used for the inhibition of enzyme activity by the phosphorylation of serine residues. Serine residues appear principally in the active center of hydrolases, such as trypsin, chymotrypsin, elastase, thrombin, and subtilisin, or in phosphatases and cholinesterases (see Table I).

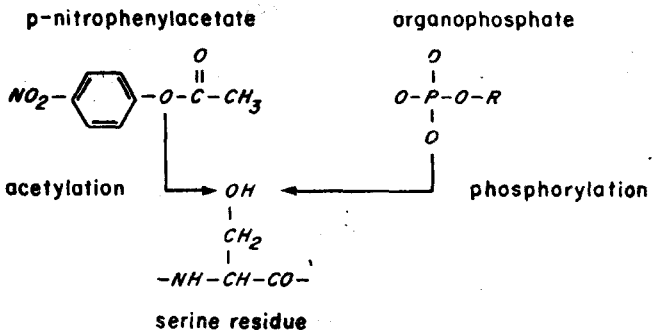
The action of hydrolases having serine as an active element proceeds along similar lines. Amide and ester moieties are cleaved and the



acyl residue is transferred to serine. This residue can then become hydrolyzed or transferred to other groups, such as the alcohols. By 1934 the hydrolysis mechanism had already been elucidated by ^{18}O-isotope techniques. It was shown that hydrolysis of acetic acid *n*-amylester in the presence of $H_2{}^{18}O$ did not yield ^{18}O-enriched *n*-amyl alcohol. The hydrolytic cleavage reaction involves an opening of the C—O bond rather than of the O—R′ bond (7):



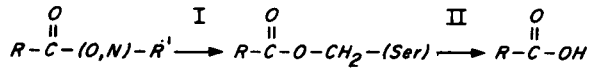
It has been shown frequently in work with ^{32}P-labeled organophosphate compounds that 1 mole of the enzyme reacts with 1 mole of the inhibitor, with a loss of the activity. Degradation studies on this enzyme revealed that the esterification of the phosphate proceeds at only one specific serine site (10–12). The relationships established between the inhibition of the active enzyme center by phosphates and the acetylation of this specific serine during the catalytic hydrolysis of *p*-nitrophenyl acetate were the final clues in the discovery of serine as the active center in hydrolases.



Inhibition of enzymes by phosphate groups is generally a reversible process; hydrolases become reactivated the moment the PO_4 group is released.

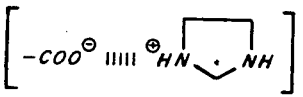
D. IONIC GROUPS AT THE ACTIVE SERINE CENTER

The cleavage mechanism for the C—O and not the O—R bond also proceeds in enzyme reactions; the acyl groups will be transferred.



Ionic groups participate in this twofold acyl transfer [eq. (1)]. Many observations suggest a participation of histidine residues in the enzymic reactions. Enzyme activity is a function of proton concentration (8); bell-shaped relationships have been noticed (Fig. 2). Dependence on the proton concentration indicates that ionic processes are involved in enzymic activities; negatively and positively charged side chain groups participate. For a review on this subject refer to reference 9.

The enzymic activity of acetylcholinesterase as a function of different H^+ concentrations (pH) has been determined (8). A maximum is established at a pH of 7.5; a 200-fold change in the H^+ ion concentration (activity) to the acid or alkaline side renders the enzyme inactive. This relationship is interpreted as resulting from the presence of an acid COO^- and a basic N^+ group in the active enzyme center, probably in the form of an ionic pair:



E. PROTEASES

Research on amino acid sequences by degradation techniques has yielded interesting results on the active serine. In the enzymes, a

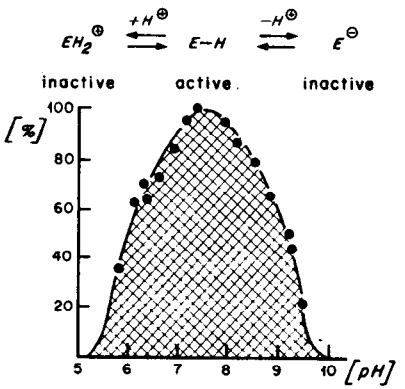


Figure 2 Rate of hydrolysis of acetylcholine by acetylcholinesterase.

distinct order of amino acids occurs next to serine. For instance, the peptide bond breaking enzymes trypsin, chymotrypsin, and thrombin are characterized by a peptide segment of the following amino acid composition:

-Gly-Asp-SER-Gly-

The presence of aspartic acid next to serine is essential for the maintenance of the enzymic activity. It comes therefore as a surprise to find in *Bacillus subtilis* a different type of peptide bond breaking enzyme, i.e., subtilisin, with the following amino acid sequence next to serine:

-Gly-Thr-SER-Met-Ala

This example represents a clear case of evolutionary development, where two types of enzymes evolve on the same substrate. A comparison of the crystallographic structures of the substrate center of these two enzyme types, in which serine is part of the active site, would have great scientific merit. The two sequences in question, Asp-SER-Gly and Thr-SER-Met, are remarkably different from one another. An accidental development of the sequence Thr-SER-Met can be ruled out for one principal reason—the protease of the mold *Aspergillus* (*a*) and the caseinase of the French bean (*b*) have the same serine neighbors as does subtilisin (*c*):

-Gly-Thr-SER-Met-Ala; (*a*) Thr-SER-Met-Ala; (*b*) Gly-Thr-SER-Met-Ala (*c*)

Much remains to be learned before the phylogenetic meaning of this phenomenon is clarified.

Although amino acids next to the active serine -Gly-Asp-SER-Gly- are identical for chymotrypsin, trypsin, elastase, and thrombin, there are considerable differences in the peptide bonds that are cleaved (13). Without question, trypsin is the most specific enzyme because it hydrolyzes only the peptide bond at the carboxyl side right of lysine and arginine: