ARTICLE

Analogous Salicylic Acid Affinity Regions in Serum Albumin and Soybean β-Conglycinin

John A. Rothfus[1] and Walter J. Wolf*

Biomaterials Processing Research and Plant Polymer Research, NCAUR, USDA, ARS, Peoria, Illinois 61604

ABSTRACT: The known structures of human serum albumin and its binding sites for salicylic acid (SA) were used as a model for examining the amino acid sequence of soybean β-conglycinin to determine whether a plant protein might share structural similarities with albumin relative to the binding of SA. Molecular mechanics energy calculations for computed nonionized and ionized interactions *in vacuo* identified the two major SA affinity regions in human serum albumin and a single analogous region in the α-subunit of soybean β-conglycinin. SA interactions with 21-residue segments from both proteins suggest redundant binding sites with multiple degrees of affinity. High-affinity segments incorporate hydrophobic amino acids with combinations or multiple residues of histidine, tryptophan, lysine, or arginine in keeping with a preference of SA for homopolymers of these acids over other homopolypeptides. Residue spacing also seems important. High affinity is associated with but not imparted by genetic similarity. Profiles from β-sheet conformations simplify identification of analogous segments.

JAOCS 74, 1193–1201 (1997).

The fundamental molecular unity of life allows convenient synergy between widely divergent species. Mammals thus derive sustenance from plants, humans armed with animal proteins can tolerate life-threatening conditions, and the traumatized have been saved by extending vital fluids with microbial polysaccharides. Molecular substitutions and augmentations sustain and enrich human existence in multiple ways yet the development of useful alternatives and amendments is often more serendipitous than systematic. Only recently, has accumulated knowledge disclosed how the evolution of protein structural domains can conserve similar function in otherwise quite dissimilar protein sequences (1). This insight provides both justification and direction for systematic analyses of plant protein sequences in programs that seek new uses for some of the world's most abundant and least expensive commodities.

In terms of adding value by substitution, immunoglobulins, hormones, enzymes, and certain other proteins from animals and recombinant microorganisms already find acceptance in medicine (2), but seed proteins remain relatively unexamined for physiological utility other than nutrition. This neglect is clearly due, in part, to inadequate technology. Detailed knowledge of plant protein structures and methods for associating specific structures with useful properties are both limited. An equally serious deterrent is the extensive immunology that derives from Wells and Osborne's administration (3) of seed proteins to animals at the turn of the century. This early work, which associated certain seed proteins with anaphylaxis, provided ample reason for caution. Ironically, it also pointed to the ultimate solution to protein allergen problems. Wells and Osborne (3) concluded that shared reactive structures are the source of shared antigenicity. Such domains now can be eliminated by breeding or they can be disrupted by site-directed mutagenesis. However, as recent expression of a Brazil nut allergen in transgenic soybeans illustrates (4), simple methods are needed to identify useful seed protein domains and segregate them from potentially deleterious segments.

Where protein sequences are known, computational techniques allow rapid prediction of conformations (5) and, thereby, estimates of molecular properties for soluble proteins. For proteins of diverse origin, simple comparisons of amino acid distributions or volume, amphiphilicity, and charge profiles of the proteins in hypothetical β-sheet conformations (6) provide quick impressions of potential functionality. As exemplified herein, such simple methods have proven advantageous in scanning seed proteins for functional structures. This work utilized knowledge of serum albumin (7,8), the major transport protein in blood, to locate a prospective small-molecule binding site in soybean β-conglycinin, a major soybean storage protein. Salicylic acid (SA) was chosen as the model ligand because its binding sites in albumin are known (8).

EXPERIMENTAL PROCEDURES

Functional comparisons were based on the primary structures of human serum albumin (HSALB) (8) and the α subunit of soybean β-conglycinin (SABCNGLN) (Fig. 1) (9).

Computerized models of selected polypeptide segments were constructed with the July 94 version of the Chem-X molecular modeling system, developed and distributed by

[1]Deceased September 10, 1996.

*To whom correspondence should be addressed at NCAUR, 1815 North University St., Peoria, IL 61604.

E-mail: wolfwj@mail.ncaur.usda.gov.

A

```
DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV KLVNEVTEFA KTCVADESAE     60
NCDKSLHTLF GDKLCTVATL RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV    120
DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR YKAAFTECCQ AADKAACLLP    180
KLDELRDEGK ASSAKQRLKC ASLQKFGERA FKAWAVARLS QRFPKAEFAE VSKLVTDLTK    240
VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE KPLLEKSHCI AEVENDEMPA    300
DLPSLAADFV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVVLLLRLA KTYETTLEKC    360
CAAHDPHECY AKVFDEFKPL VEEPQNLIKQ NCELFKQLGE YKFQNALLVR YTKKVPQVST    420
PTLVEVSRNL GKVGSKCCKH PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES    480
LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKE RQIKKQTALV ELVKHKPKAT    540
KEQLKAVMDD FAAFVEKCCK ADDKETCFAE EGKKLVAASQ AALGL
```

B

```
VEKEECEEGE IPRPRPRPQH PEREPQQPGE KEEDEDEQPR PIPFPRPQPR QEEEHEQREE     60
QEWPRKEEKR GEKGSEEEDE DEDEEQDERQ FPFPRPPHQK EERNEEEDED EEQQRESEES    120
EDSELRRHKN KNPFLFGSNR FETLFKNQYG RIRVLQRFNQ RSPQLQNLRD YRILEFNSKP    180
NTLLLPNHAD ADYLIVILNG TAILSLVNND DRDSYRLQSG DALRVPSGTT YYVVNPDNNE    240
NLRLITLAIP VNKPGRFESF FLSSTEAQQS YLQGFSRNIL EASYDTKFEE INKVLFSREE    300
GQQQFEQRLQ ESVIVEISKE QIRALSKRAK SSSRKTISSE DKPFNLRSRD PIYSNKLGKF    360
FEITPEKNPQ LRDLDIFLSI VDMNEGALLL PHFNSKAIVI LVINEGDANI ELVGLKEQQQ    420
EQQQEEQPLE VRKYRAELSE QDIFVIPAGY PVVVNATSNL NFFAIGINAE NNQRNFLAGS    480
QDNVISQIPS QVQELAFPGS AQAVEKLLKN QRESYFVDAQ PKKKEEGNKG RKGPLSSILR    540
AFY
```

FIG. 1. In single-letter notation, amino acid sequences: A: human serum albumin, B: soybean β-conglycinin (α subunit).

Chemical Design Ltd., Oxford, England (Chemical Design Inc., Mahwah, NJ), and examined within the Chem-X system for ability to form stable associations with SA either ionized (SAI) or nonionized (SAN). With the Chem-X system, complexes were easily minimized in terms of geometry and van der Waals interaction energy (VDW) or optimized in terms of molecular mechanics energy (MME). Computations were performed on a Digital MicroVAX 3100 computer (Digital Equipment Corp., Marlborough, MA) equipped with a Tektronix 4225 high-resolution terminal (Tektronix Inc., Beaverton, OR).

Prior to addition of the SA guest molecule, peptide segments constructed as random coils were converted to extended chains with *trans* backbone torsion angles and then optimized in terms of molecular mechanics energy through 100 cycles or to convergence (<0.1 calorie change per cycle). Proline residues, if present, produced *trans*-structure deviations that were left unchanged. The optimized extended chain served as a standard, even though arbitrary, starting conformation for subsequent minimization and association with SA.

Electron distributions in the protein segments and SA were set according to Gasteiger and Marsili (10) for computations involving SAN complexes, or they were augmented in SAI complexes with additional specific atomic charges as follows: –0.5 on OD_1 and OD_2, the side-chain oxygens of aspartic acid (Asp); –0.5 on OE_1 and OE_2 of glutamic acid (Glu); 1.0 on NZ, the side-chain nitrogen of lysine (Lys); 0.5 on NH_1 and NH_2, guanido nitrogens of arginine (Arg). Favored peptide conformations were obtained by incrementing backbone torsion angles and monitoring van der Waals energy to convergence or through 100 cycles. This minimization frequently produced curved segments with terminal N and C atoms separated, in some instances, by as little as 10 Å.

Guest–host affinity was assayed by placing an optimized SA molecule, SAI or SAN, into the plane defined approximately by the center of a curved peptide segment and its terminal residues. Carboxyl oxygens of SAI were differentiated by unequal charge distribution: –0.2 on the carbonyl oxygen (O_{22}), and –0.8 on the carboxylate oxygen (O_{20}). MME and VDW for optimized SAN were 14.0 and 6.6 kcal; for SAI, 0.8 and 9.2. The guest molecule position was adjusted to maximize within-range atomic interactions, to minimize atomic pairs in van der Waals interaction, and to eliminate invalid interactions, which tended to cause guest–host separation. Generally, minimization started with 4–6 carbonyl carbons in the peptide backbone less than 10 Å from C_1 of the SA aromatic ring. Computations were performed in total energy mode measuring interactions between the guest and all protein atoms within a range equal to at least half the peptide segment length: i.e., usually 1 to 40 Å for a 21-residue peptide, with H-bond = 1.31, ratio = 0.6, and dielectric constant = 1.0. Minimization for 100 cycles established atom pair interactions that controlled binding. Subsequent optimization for 100 cycles determined the ultimate energy level of the guest–host complex. Measuring the stability of each complex in terms of the total energy for SA atom interactions with all others in the complex allowed comparison of SA binding to peptides of different size and stability. The energy of complex formation

(ΔMME) was approximated for comparison purposes by subtracting the sum of the MME for the uncombined guest and host from the total MME for the guest–host complex. This affinity assay was generally conducted at least twice with each segment. Three or more probes with the guest in different positions were common when primary structure varied substantially within a segment.

Profiles of molecular volume, amphiphilic character, and ionized residue distribution for β-sheet proteins and peptides were computed from amino acid sequences by the moving window analysis of Rose *et al.* (11) modified to segregate odd- and even-residue properties and to measure peptide volumes in cubic angstroms ($Å^3$) and amphiphilicity in arbitrary units (a.u.) (6). Analogies between proteins were identified by visual inspection of superpositioned profiles. Genetic codon similarities were identified by automated matching of polypeptides in both parallel (N-terminus over N-terminus) and antiparallel (N-terminus over C-terminus) orientations. For purposes of this analysis, similar codons were either identical or within one nucleic acid replacement of identity. Significant similarities or differences deviated by at least one standard deviation from averages established for each peptide chainlength by analyses of random sequences generated from a randomized amino acid population weighted according to the frequency of amino acid occurrence (12). For example, a high level of similarity for a 21-residue peptide would exceed 70%. Average for this chainlength is 64%. These values increase and decrease, respectively, for shorter or longer fragments. For a 47-residue peptide, average is 57%, and a high level of similarity would exceed 60%.

RESULTS AND DISCUSSION

Guest binding to a host protein is essentially a time–space process controlled by geometry and the relative strengths of competing interactions. Various initial interactions can lead to essentially equivalent minimal energy states, given enough time. Complexes that form quickly and are relatively more stable than a mixture of unbound molecules thus seem most consistent with the transport function of proteins like HSALB. Accordingly, each probe of a particular peptide host with SA was limited to a 100-cycle minimization followed by a 100-cycle optimization. This arbitrary protocol proved a convenient alternative to time-prohibited generation of the global energy surface for each SA–peptide complex. Interactions that ultimately led to low-energy conformations in random lengthy (1000–3000 cycles) convergence minimizations often exhibited rapid energy decline during the first 60 cycles of minimization and less than 1% change per cycle after 90 cycles. Global–minimal complexes, which might form only after prolonged perturbation, were therefore assumed to be represented among those identified by the protocol.

Homopeptide binding. Representative 21-residue homopolypeptides were probed with SA to identify types and relative strengths of associations and to provide a frame of reference within which sequence–specific interactions might be detected. With nonionized guest and host, electrons were distributed according to Gasteiger and Marsili (10). For probes involving SAI, however, specific charges were also set on ionizable carboxyl, amino, or guanido groups, and the charged peptides were minimized prior to interaction with SAI.

TABLE 1
Salicylic Acid–Homopolypeptide Interactions[a]

Amino	Nonionized (SAN)			Ionized (SAI)			
acid	MME	VDW	ΔMME[b]	MME	VDW	ΔMME[b]	Contacts[c]
None	14	7		1	9		
Arg	6	–10	–31	–36	–43	–75	1
				–69	–61	–107	3
Lys	9	–5	–28	–143	–87	–172	1
				–190	–180	–202	2
				–278	–209	–297	3
Glu	–11	–27	–46	–3	–11	–23	1
				–10	–18	–27	2
His	–15	–30	–55	–223	–231	–233	5
				–295	–300	–299	9
Trp	21	7	–41	–134	–144	–138	4
Leu	13	–4	–22	10	–1	–7	
Phe	13	–3	–23	9	0	–5	
Tyr	14	–2	–21	3	–7	–9	
Gln	17	1	–21	11	3	–5	
Pro	15	–1	–28	5	–5	–11	
Ser	16	1	–20	9	1	–5	

[a]Molecular mechanics (MME) and van der Waals (VDW) energies (kcal) of salicylic acid atom interactions with host polypeptide atoms.
[b]Total energy (kcal) of guest–host complex minus the sum of energies for the uncombined guest and host. Arg, arginine; Lys, lysine; Glu, glutamic acid; His, histidine; Trp, tryptophan; Leu, leucine; Phe, phenylalanine; Tyr, tyrosine; Gln, glutamine; Pro, proline; Ser, serine.
[c]Number of amino acid residues with atoms less than 4 Å from guest atoms.

Results given in Table 1 show that under conditions of these tests SA forms the most stable complexes with polar amino acids or those with heterocyclic nitrogen that can delocalize charge easily. Stability correlates with negative energy values and the magnitude of energy change (ΔMME) upon complex formation. As expected, ionization weakens SAI association with acidic amino acids but strengthens binding to bases. With charged poly-Lys, stability increases in proportion to the number of amino groups associated with the carboxylate group of SAI. SA ionization also destabilizes nonpolar interaction with neutral and hydrophobic amino acids. Poly-histidine (His) and poly-tryptophan (Trp) form some of the most stable complexes with either SAN or SAI.

This SA affinity for poly-His and poly-Trp is interesting because the single Trp in HSALB occurs in one of the protein's two SA binding sites (8). Analogous binding at the second site thus requires an equivalent surface constituted without Trp. The results suggest that one or more His residues or a combination of basic amino acids might easily allow similar affinity.

Even though binding tended to disrupt secondary structure, SA binding to homopolypeptides was not clearly independent of backbone conformation. Those peptides that exhibited the greatest affinity were frequently helical after minimization and before contact with SA. Poly-Lys, -Glu, -His, and -Trp all minimized to helical structures with 4–5 residues per turn. The effect of this helical structure was evident in a tendency of poly-Lys to form multidentate coordination through the ε-amino group of every fourth or fifth residue rather than those of adjacent lysines. Poly-proline (Pro) was equally helical, but a relatively poor binder. Homopeptides in which β-sheet character predominated [poly-leucine (Leu), -phenylalanine (Phe), -tyrosine (Tyr), and -serine (Ser)] also exhibited poor affinity. Poly-Arg, which minimized to a β-sheet when nonionized or a helix with slightly more than three residues per turn when ionized, exhibited intermediate affinities. Poly-glutamine (Gln), another intermediate helix with slightly more than three residues per turn, bound SA rather poorly.

Serum albumin. Site I, at which HSALB binds aromatic guest molecules, extends approximately from residue 180 through 290 (8). In terms of SA binding, it is significant that 20 of 83 Arg and Lys residues in HSALB are concentrated in this segment along with the molecule's single Trp and three of its 16 His residues. Several 110-residue segments between residues 105 and 253 match this concentration of highly interactive amino acids, and all of them contain Trp. Two other 110-residue segments, located between residues 432 and 545, contain 19 basic amino acids and four His but no Trp.

Initial probes of 100-residue HSALB segments confirmed SA binding capability. For example, segment (191–291) bound SAN with MME = −7 kcal and ΔMME = −21 kcal. These trials also identified an interesting procedural complication. When perturbed by SA insertion, lengthy peptides frequently minimized to conformations that excluded SA. This was especially noticeable in SAN probes of nonionized fragments, where binding depended on relatively weak associative forces. Apparently in such cases, there is greater chance that bond rotations inherent in minimization will separate the guest and host before any energetic benefit of interaction becomes sufficient to impact the minimization process. The most effective interactions were generally established early during minimization.

SA also often interacted with lengthy peptides through single residues, and such complexes frequently stayed in local energy minima rather than transform to global–minimal conformations with multidentate interactions. These results, and the concentration of interactive amino acids around Trp, suggest that the protein's observed binding characteristics probably represent a composite of associations that are qualitatively similar but quantitatively different. Such an ability to bind with an array of affinities should impart a valuable carrier protein property: i.e., the ability to transfer ligands to a variety of alternative hosts.

Analysis of shorter segments with fewer atom-atom interactions reduced computation times yet still differentiated satisfactorily between degrees of stability. With shorter sequences, the energetics of SA interaction were affected less by unbound portions of each protein segment, and MME and ΔMME values increased in negative magnitude. SAN expulsion decreased, and preferred sites were still detected. For example, values from repeated probes of a 47-residue peptide (HSALB 180–226) were lowest when either SAN or SAI was bound near Trp (214) with the carbonyl carbon associated with the ε-amino group of Lys (205). With SAI, the amino group of Lys (181) also associated frequently with the carboxylate group. MME and ΔMME for SAN and SAI binding to HSALB (180–226) were 9 and −29, and −176 and −198, respectively. This preference for the region around Trp (214) is consistent with other studies (8,13,14).

Some of the lowest SA energy levels were achieved with HSALB (203–223), which contains, in addition to Trp, five basic amino acids that provide for multiple interactions with SA. Fragments shorter than 20 residues tended to produce less-stable complexes in which SA was associated with a single amino or guanido group.

Table 2 summarizes data from probes of HSALB fragments with SAN and SAI. The analyses show considerable variation in MME and ΔMME values throughout the length of HSALB. MME reflects the relative stability of SAN or SAI within the environment of the peptide while ΔMME measures energy change due to complex formation. In general, these two quantities support one another: i.e., lower MME values for bound SA agree with larger changes in the MME of the two-molecule system upon complex formation. It is difficult to attach significance to small differences in ΔMME because the data do not differentiate clearly between ΔMME due to binding and ΔMME enabled by binding. SA binding may activate a segment sufficiently to allow shifts to lower energy conformations in portions of the peptide that are not critical to binding. Random extended analyses, for example, found that an added 100-cycle optimization of the minimized starting peptide reduced the magnitude of ΔMME values 10–15%,

TABLE 2
Serum Albumin–Salicylic Acid Interactions[a]

Segment[b]	Sequence	Nonionized (SAN) MME	Nonionized (SAN) ΔMME[c]	Ionized (SAI) MME	Ionized (SAI) ΔMME[c]
120–140	VDVMCTAF**H**DNEETFL**KK**YLY	7	–34	–40	–54
138–158	YLYEIA**RRH**PYFYAPELLFFA	2	–36	–20	–39
165–185	FTECCQAAD**K**AACLLP**K**LDEL	4	–41	–100	–121
189–209	G**K**ASSA**K**Q**R**L**K**CASLQ**K**FGE**R**	–1	–36	–115	–133
203–223	LQ**K**FGE**R**AF**K**A**W**AVA**R**LSQ**R**F	–9	–38	–124	–136
238–258	LT**K**V**H**TECC**H**GDLLECADD**R**A	–6	–47	–93	–115
280–300	E**K**PLLE**K**S**H**CIAEVENDEMPA	–2	–39	–108	–121
327–347	LGMFLYEYA**RRH**PDYSVVLLL	4	–33	–39	–56
361–381	CAAADP**H**ECYA**K**VFDEF**K**PLV	14	–29	–67	–86
401–421	Y**K**FQNALLV**R**YT**KK**VPQVSTP	10	–25	–110	–118
420–440	TPTLVEVS**R**NLG**K**VGS**K**CC**KH**	9	–28	–104	–122
426–446	VS**R**NLG**K**VGS**K**CC**KH**PEA**KR**M	–6	–40	–140	–157
460–480	LCVL**H**E**K**TPVSD**R**VT**K**CCTES	13	–26	–50	–66
488–508	FSALEVDETYVP**K**EFNAETFT	14	–27	–88	–105
499–519	P**K**EFNAETFTF**H**ADICTLSE**K**	4	–41	–49	–69
518–538	E**K**E**R**QI**KK**QTALVELV**KHK**P**K**	10	–36	–105	–127
180–226		9	–29	–176	–198
190–219		3	–52	–158	–178
201–250		5	–38	–88	–114

[a]Molecular mechanics (MME) energies (kcal) of salicylic acid atom interactions with host polypeptide atoms.
[b]Inclusive residue numbers.
[c]Total energy (kcal) of guest–host complex minus the sum of energies for the uncombined guest and host.

hardly enough to merit additional computing time. Energy values produced by these analyses are thus better suited for comparison than absolute characterization.

Consistent with the affinities of homopeptides, segments that contain relatively high concentrations of Lys, Arg, His, and Trp produce large ΔMME. Several such segments occur in Site I, which includes HSALB (203–223). Similar high-affinity segments occur in the region centered about HSALB (426–446), which lies within the second aromatic binding site identified by Carter and Ho (8). Others, notably HSALB (165–185), (280–300), (420–440), and (518–538), bind SAN and SAI somewhat less avidly but with nearly equal affinities even though constituted with quite different combinations of high-affinity acids and other types. This evidence of sequence specificity is further corroborated by substantial differences between segments in which high-affinity residues are clustered, HSALB (138–158) and (327–347), vs. others in which they are dispersed.

Soybean β-conglycinin. SABCNGLN differs distinctly from HSALB and other proteins due to an exceptionally high concentration of polar amino acids near its N-terminus. Specifically, 48% of the molecule's acidic amino acids and 27% of its basic amino acids occur in the N-terminal 22% of the sequence. The remainder of the molecule curiously resembles portions of HSALB when the two proteins are compared in profile (Fig. 2).

Beyond residue 120 of SABCNGLN, segments in eight different locations, near residues 140, 175, 235, 285, 310, 365, 480, and 510, are analogous to HSALB (203–223) in terms of volume, amphiphilicity, and ionizable residue profiles. All, however, are constituted differently and exhibit different affinities for SAN and SAI. Table 3 lists MME and ΔMME values for 21-residue fragments from throughout the SABCNGLN sequence.

Three of these analogous locations i.e., near residues 135, 330 and 525, concentrate high-affinity amino acids and should therefore bind SA strongly. Indeed, segments (317–337), (341–361), and (521–541), which average more than 30 residue percentage Arg plus Lys, complex SAI with ΔMME values among those that best duplicate the value for HSALB (203–223). However, none of these segments contains His or Trp. A nearby fragment, (382–402), and another, (178–198) each contains only one His and one Lys, and both bind SA with ΔMME of notable magnitude. Three additional His-containing segments, (10–30), (87–107) and (126–146), produce intermediate-level ΔMME with different combinations of amino acid type and concentration.

Among the least-stable complexes with segments containing high-affinity residues are two that represent a fivefold variation in peptide polarity. Segment (104–124) contains one Arg but is overwhelmingly acidic (66 residue percentage Asp plus Glu) with overall amphiphilicity of 1354 a.u. Segment (257–277) also contains one Arg and two acidic residues but is otherwise relatively hydrophobic (262 a.u.). Since these two bind SA with about the same ΔMME, it appears that segment amphiphilicity is not critical even though it can contribute to complex stability.

Essentially equivalent binding can be achieved in various

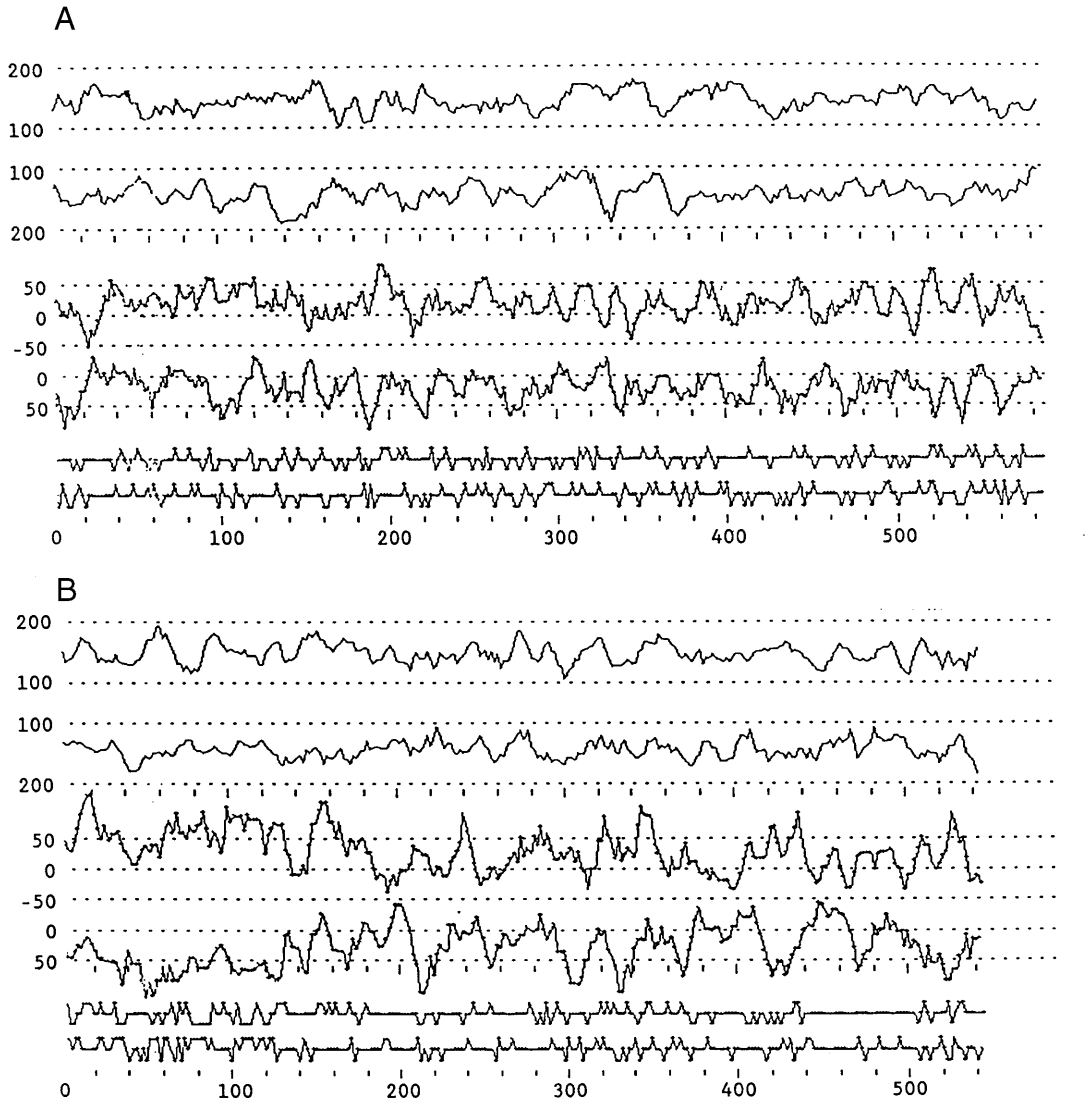

FIG. 2. Volume, amphiphilicity, and ionizable residue profiles of (A) human serum albumin, (HSALB), and (B) soybean β-conglycinin (α-subunit), (SABCNGLN). Upper pair of curves traces volume in $Å^3$; middle, amphiphilicity in arbitrary units; lower, ionizable residue location (negative charges can occur at indentations along the profiles; positive charges at protrusions). Upper curve in each pair represents Side A of the peptide chain; lower, Side B. Properties are displayed, left to right, from N-terminus to C-terminus. Tick marks indicate 20 residues.

ways. High-affinity residues ensure a certain stability, but differences in formation energies (Table 3) suggest levels of composition and sequence control that are not yet obvious from the data. For example, SABCNGLN (476–496), which contains no high-affinity residues and is less hydrophilic than HSALB (203–223), i.e., 72 vs. 478 a.u., expectedly exhibits poor affinity for SA, but a nearby segment, SABCNGLN (499–519), is not much better even though it contains three basic residues and is amphiphilically equivalent to the albumin peptide. The spacing of high-affinity residues relative to hydrophobic residues could be important, but additional work is needed to determine exactly how sequence variation regulates binding.

SABCNGLN (50–70) best illustrates the impact of high-affinity residues and the importance of nonpolar acids. This segment rivals HSALB (203–223) in terms of Arg, Lys, His, and Trp content and in terms of affinity for SAN and SAI. It is appealing to speculate that a polypeptide so constituted with many ionizable acidic residues might readily bind and solubilize SA or other aromatic acids. The data indicate that SAN would bind strongly to segment (50–70), but suggest that ionized complexes might be slightly less stable than those

TABLE 3
Soybean β-Conglycinin–Salicylic Acid Interactions

Segment	Sequence	Nonionized (SAN) MME	Nonionized (SAN) ΔMME	Ionized (SAI) MME	Ionized (SAI) ΔMME
10–30	EIP**R**P**R**P**R**PQ**H**PE**R**EPQQPGE	16	–29	–55	–77
50–70	**R**QEEE**H**E**WR**EEQE**W**P**RK**EE**KR**	11	–32	–103	–124
87–107	DE**R**QFPFP**R**PP**H**Q**K**EE**R**NEEE	4	–34	–60	–77
104–124	NEEEDEDEEQQ**R**ESEESEDSE	–2	–32	–11	–29
126–146	**RRHK**N**K**NPFLFGSN**R**FETLF**K**	12	–26	–73	–92
134–154	FLFGSN**R**FETLF**K**NQYG**R**I**R**V	10	–47	–84	–101
141–161	FETLF**K**NQYG**R**I**R**VLQ**R**FNQ**R**	3	–35	–87	–102
167–187	NL**R**DU**R**ILEFNS**K**PNTLLLPN	11	–28	–57	–77
178–198	S**K**PNTLLLPN**H**ADADYLIVIL	3	–38	–103	–117
190–210	DADYLIVILNGTAILSLVNND	14	–25	8	–12
224–244	RVPSGTTYYVVNPDNNENLRI	13	–27	–33	–50
235–255	NPDNNENL**R**LITLAIPVNKPG	13	–26	–71	–99
257–277	FESFFLSSTEAQQSYLQGFS**R**	–7	–29	–10	–28
268–288	QQSYLQGFS**R**NILEASYDT**K**F	14	–26	15	–5
275–295	FS**R**NILEASYDT**K**FEEIN**K**VL	10	–30	–106	–125
302–322	QQQGEQ**R**LQESVIVEIS**K**EQI	3	–31	–44	–57
305–325	GEQ**R**LQESVIVEIS**K**EQI**R**AL	12	–27	–73	–87
312–332	SVIVEIS**K**EQI**R**ALS**KR**A**K**SS	5	–31	–109	–125
317–337	IS**K**EQI**R**ALS**KR**A**K**SSS**RK**TI	12	–23	–116	–133
328–348	**R**A**K**SSS**RK**TISSED**K**PFNL**R**S	7	–31	–114	–130
341–361	D**K**PFNL**RSR**DPIYSN**K**LG**K**FF	13	–23	–116	–127
362–382	EITPE**K**NPQL**R**DLDIFLSIVD	5	–42	–93	–113
382–402	DMNEGALLLP**H**FNS**K**AIVILV	3	–35	–89	–102
476–496	FLAGSQDNVISQIPSQVQELA	12	–27	3	–14
499–519	GSAQAVE**K**LL**K**NQ**R**ESYFVDA	6	–32	–8	–27
521–541	P**KKK**EEGN**K**G**RK**GPLSSIL**R**A	–5	–41	–89	–115
315–361		0	–19	–180	–203
318–347		7	–33	–147	–170
510–535		6	–35	–84	–105

of SAI with HSALB (203–223) or SABCNGLN (317–337). Such diminished stability would be consistent with the fact that SABCNGLN (50–70) is more polar (1593 a.u.) than the albumin fragment (478 a.u.) or the other soy fragment (925 a.u.). In this regard, it is interesting to note that ΔMME of −36 and −171 were obtained for SAN and SAI complexes with a hypothetical analog of SABCNGLN (50–70), RQEFEHVWRFLQLWPRKFEKR, in which six glutamic acid residues were replaced by two Leu, three Phe and one valine (Val): i.e., in proportions at which they occur in HSALB (203–223). Another analog, RQEVEHVWRVVQVWPRKVEKR, in which six Val replace Glu, is genetically more feasible because conversion of Glu to Val requires only a single nucleic acid base change in each altered codon. This simpler peptide bound SAN and SAI with energies of –40 and –112 kcal, respectively.

The importance of segment composition and sequence is further demonstrated by SABCNGLN (126–146) and (134–154). Both of these segments concentrate basic amino acids to nearly the same extent as those that exhibit the greatest affinities for SA, but they both bind SAI with less affinity than, for example, does 312–332 or 328–348.

Some SABCNGLN segments that are associated with SA affinity also exhibit more than 70% genetic codon similarity to HSALB (203–223). For example, SABCNGLN (312–332), which is 76% similar, binds SAI with relatively high affinity (ΔMME −125). An overlapping peptide, SABCNGLN (317–337), is not so similar (71%), but it exhibits the greatest affinity among examined short segments (ΔMME −133). Interestingly, these two soy peptides share only one identical residue with HSALB (203–223): alanine (Ala), HSALB (215), SABCNGLN (324). This may be an example of genetic redundancy conserving redundant functionality.

Nineteen segments in SABCNGLN exceed 75% codon similarity to HSALB (203–223) and HSALB (426–446) (Table 4). The 12 soy segments related to HSALB (203–223) duplicate residues at 16 of the 21 positions in the serum peptide. Leu, HSALB (203), is duplicated most frequently; it appears in four of the 12 fragments. However, Phe at 206 and 211, Trp at 214, Ala at 217, and Ser at 220 in the serum protein are not duplicated in any of the soy fragments. Also, replacements at the Phe and Trp locations frequently require multiple base changes in the residue codon: i.e., these residues are removed farthest genetically from counterparts in HSALB (203–223). Conversely, Arg at 209 and 218 in the serum protein are either duplicated in the soy fragments or replacements require no more

TABLE 4
β-Conglycinin Sequences Genetically Similar to High-Affinity Peptides of Albumin

Segment	Sequence[a]	Similarity[b]
HSALB		
203–223	**L Q K F G E R A F K A W A V A R L S Q R F**	
SABCNGLN		
37–57R[c]	**W E** H E **E E Q** R P **Q P R P F P I P R P Q** E	76
41–61R	**Q E E** R **W E H E** E E Q **R P** Q **P R P F P I** P	76
244–264	**L** I **T L A** I **P V** N **K P G** R **F E S F F L S S**	81
302–322	**Q Q Q G E Q R** L Q **E S** V I **V E I S** K E Q I	76
305–325R	**L** A **R I** Q **E K S I E V** I V S **E Q L R Q** E **G**	76
312–332	**S** V **I V E** I **S** K E **Q** I **R A L S K R A K S S**	76
328–348	**R** A **K S S** S **R** K T **I S S E D** K **P F N L R S**	76
369–389R	**L L** A **G E** N **M D V I S L** F **I D L** D **R L Q** P	76
395–415	**S K** A **I V** I **L V I N E G D A** N **I** E **L** V G L	76
398–418R	**Q E K L G V L E I N A** D **G E** N **I V L** I V **I**	81
476–496R	A **L E** Q **V Q S P I Q S** I V N **D Q S G** A **L F**	76
488–508R	**L L K** E **V A Q A S** G **P** F **A L E Q V** Q S **P I**	76
HSALB		
426–446	**V S R N L G K V G S K C C K H P E A K R M**	
SABCNGLN		
146–166	K **N Q Y** G **R I** R **V L Q R F N Q R** S **P Q L** Q	76
150–170R	**D R L N Q** L **Q** P **S R Q** N **F R Q L V** R **I R** G	76
289–309	**E** E **I N** K **V** L **F S R E** E **G Q Q Q G E Q R L**	81
309–329R	**A R K S L A R I** Q E **K S** I **E** V I **V S E Q L**	76
324–344	**A L S K R A K** S **S S R** K T **I** S **S E D K P** F	76
423–443R	**I** D **Q** E **S** L **E A R Y K R** V **E L P Q E E Q** Q	76
488–508	**I P S** Q **V** Q **E L A F** P **G S** A **Q A V E K L L**	81

[a]Similar residues are in bold type.
[b]Percentage of codons within one nucleotide of identity with albumin codon.
[c]R indicates sequence is given in reverse orientation, C –> N, HSALB, human serum albumin; SABCNGLN, soybean β-conglycinin.

than a single base change. Why basic residues should be conserved at these two positions is not obvious. High-affinity segments SABCNGLN (312–332) and (328–348) show that the Arg are not critical to SA binding, but basic amino acids located in a relatively hydrophobic environment and separated by 7–8 residues may be. SABCNGLN (476–496), which contains no basic acids, and SABCNGLN (305–325), which has three basic acids separated by 3 and 10 residues, are both less effective binders than (312–332), (317–337), and (328–348), in which there are several ways that pairs of basic amino acids might be separated by seven residues. Interestingly, the high-affinity region in soy is also genetically similar to HSALB (426–446), which contains bases separated by seven residues in the second SA binding site of the serum protein. The HSALB sequence bounded by Arg 209 and 218 and analogous sequences from the five soy segments are all quite amphiphilically sided in β-sheet conformations. They would all be capable of presenting hydrophobic or hydrophilic surfaces to a SA guest molecule.

Time constraints confined this work to minimal-length peptides. SAI complexes with slightly longer segments, HSALB (190–219) or (180–226) and SABCNGLN (318–347) or (315–361), which include identified high-affinity sites, were also relatively stable even though residues involved in the binding (Fig. 3) were different from those in the shorter segments. Both proteins thus appear capable of redundant binding. In each, SAI associates with the ε-amino groups of two Lys and an aromatic center, either Trp in HSALB or Phe in SABCNGLN, and there is a certain geometric similarity in the two complexes. Throughout most of their sequences, however, the two peptides adopt rather different conformations. Complexes in these conformations might react differently toward other macromolecules.

The fact that these extended segments from soybean and serum continued to mimic each other suggests that even longer polypeptides or multisubunit proteins whose conformations provide essentially the same environment for SA also should be effective binders. Consistent with the high levels of affinity observed for longer segments, HSALB (180–226) and SABCNGLN (319–365) or (313–359) exhibit significant genetic similarity with 62% or more of their codons either identical or within one base replacement from identity.

These studies support the use of β-sheet profiles to identify sites of chemical analogy that could allow proteins of diverse origin to share specific functionality. While quite likely that analogous segments identified in this work will exhibit similar chemical properties, it is also quite unlikely that the whole proteins in which they occur will perform interchangeably. Nonetheless, it is intriguing to anticipate that structural elements that can mimic useful properties of less-readily-available proteins may be distributed broadly in seeds.

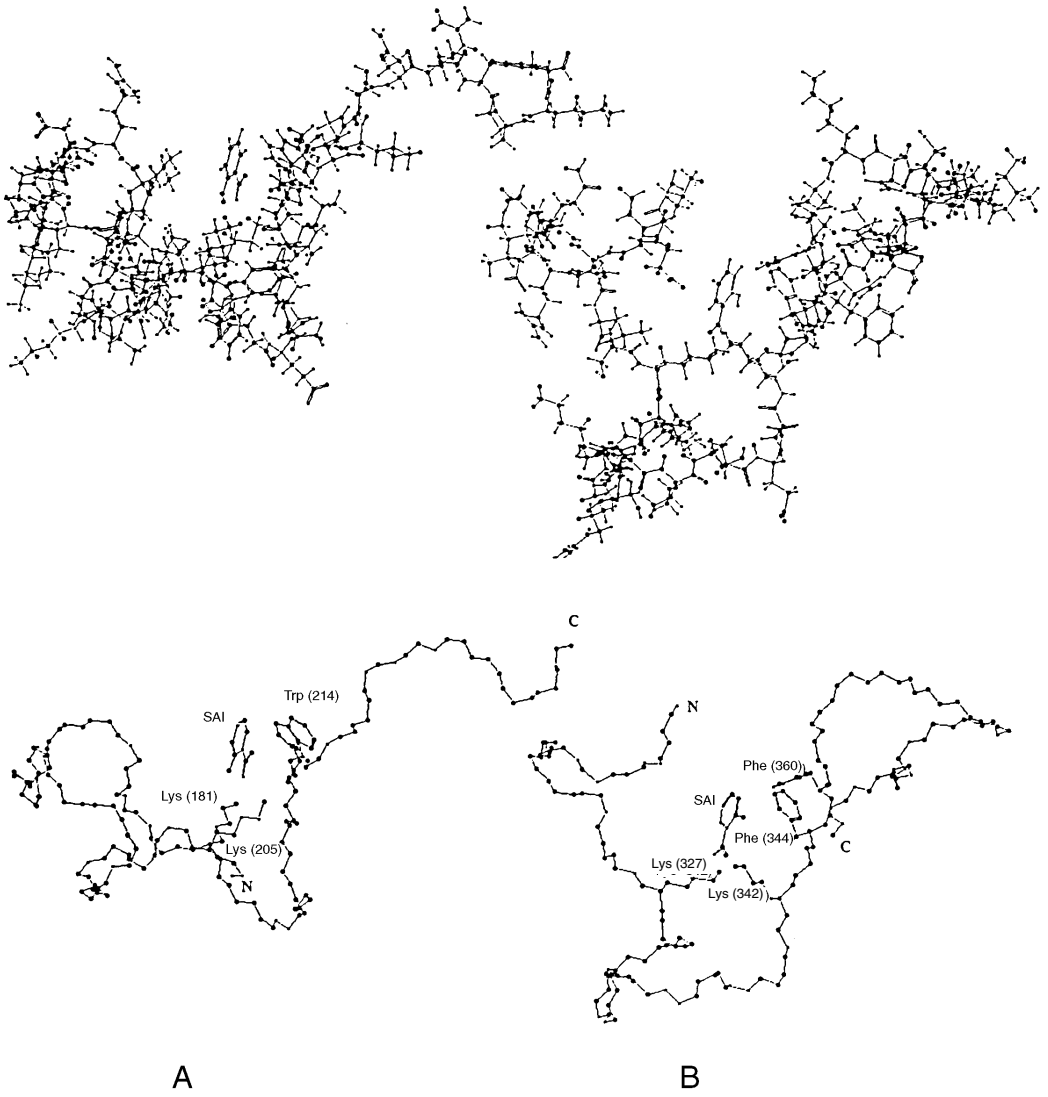

FIG. 3. Ionized salicylic acid (SAI) at prospective high-affinity sites in 47-residue peptides from (A) human serum albumin and (B) soybean β-conglycinin. Lower schematics trace the polypeptide backbone in each segment.

REFERENCES

1. Doolittle, R.F., The Multiplicity of Domains in Proteins, *Annu. Rev. Biochem. 64*:287–314 (1995).
2. Wetzel, R., Medical Applications of Protein Engineering, in *Protein Tailoring for Food and Medical Uses*, edited by R.E. Feeney and J.R. Whitaker, Marcel Dekker, Inc., New York, 1986, pp. 181–213.
3. Wells, H.G., and T.B. Osborne, The Biological Reactions of the Vegetable Proteins. I. Anaphylaxis, *J. Infect. Dis. 8*:66–124 (1910).
4. Nordlee, J.A., S.L. Taylor, J.A. Townsend, L.A. Thomas, and R.K. Bush, Identification of a Brazil-Nut Allergen in Transgenic Soybeans, *N. Eng. J. Med. 334*:688–692 (1996).
5. Srinivasan, R., and G.D. Rose, LINUS: A Hierarchic Procedure to Predict the Fold of a Protein, *Proteins: Struct. Funct. Genet. 22*:81–99 (1995).
6. Rothfus, J.A., Potential β-Sheet Surfaces of Soybean Seed Proteins, *J. Am. Oil Chem. Soc. 72*:501–506 (1995).
7. Peters, T. Jr., Serum Albumin, *Adv. Protein Chem. 37*:161–245 (1985).
8. Carter, D.C., and J.X. Ho, Structure of Serum Albumin, *Ibid. 45*:153–203 (1994).
9. Sebastiani, F.L., L.B. Farrell, M.A. Schuler, and R.N. Beachy, Complete Sequence of a cDNA of α Subunit of Soybean β-Conglycinin, *Plant Mol. Biol. 15*:197–201 (1990).
10. Gasteiger, J., and M. Marsili, Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges, *Tetrahedron 36*:3219–3228 (1980).
11. Rose, G.D., L.M. Gierasch, and J.A. Smith, Turns in Peptides and Proteins, *Adv. Protein Chem. 37*:1–109 (1985).
12. McCaldon, P., and P. Argos, Oligopeptide Biases in Protein Sequences and Their Use in Predicting Protein Coding Regions in Nucleotide Sequences, *Proteins: Struct. Funct. Genet. 4*:99–122 (1988).
13. Muralidhara, B.K., and V. Prakash, Interaction of 3′-*O*-Caffeoyl D-Quinic Acid with Human Serum Albumin, *Int. J. Peptide Protein Res. 46*:1–8 (1995).
14. Oida, T., Proton Nuclear Magnetic Resonance Study on the Multimode Interactions of Human Serum Albumin with Drug Molecules, *J. Biochem. 100*:99–113 (1986).

[Received August 7, 1996; accepted May 22, 1997]