J Mol Model (2002) 8:50–57
DOI 10.1007/s00894-001-0063-8

ORIGINAL PAPER

Rajani R. Joshi · Subramanian Jyothi

Ab initio structure of human seminal plasma prostatic inhibin gives significant insight into its biological functions

Received: 5 June 2001 / Accepted: 29 October 2001 / Published online: 12 February 2002

Abstract Human seminal plasma prostatic inhibin (HSPI) is a protein isolated from the human prostate gland. Despite its profound biomedical and biotechnological importance, the 3D structure of this protein of 94 amino acids remains undeciphered. The difficulties in extracting it in pure form and crystallizing it have restrained the determination of its structure experimentally. The homology-based computational methods are also not applicable, as HSPI lacks sufficient sequence homology with known structures in the protein data banks. We have predicted the structure of HSPI by a knowledge-based method using nonparametric multivariate statistical techniques. Stereochemical and other standard validation tests confirm this to be a well-refined structure. The biophysical properties exhibited by this structure are in good agreement with the NMR experimental observations. Docking and other computational studies on this structure provide significant explanation and insight into its binding activities and related biological and immunogenic functions and offer new directions for its potential applications.

Keywords Human seminal plasma prostatic inhibin · Ab initio structure prediction · Nonparametric regression · Protein docking · Structure–function studies

Introduction

HSPI (human seminal plasma prostatic inhibin) is a protein with 94 amino acids isolated from human prostate glands. It has been implicated in a wide range of biological activities ranging from preventing pregnancy to curing prostate cancer. It suppresses prolactin, a hormone that promotes lactation. Hence neutralizing inhibin through active immunization is seen to increase milk production. [1] To the best of our knowledge, the 3D structure of this protein has not yet been determined experimentally. Further, there is no protein with known structure in the PDB [2] that has significant (>25%) sequence identity with HSPI.

In view of its profound biological importance, the 3D structure of HSPI is long desired. Experimental determination of the structure has not been feasible because of the difficulties in getting the protein in pure form and crystallizing it. Homology modelling approaches also fail here due to lack of a suitable template.

We had predicted the 3D structure of HSPI by our nonparametric regression approach [3] using only its primary sequence (c.f. [4] or Swiss-Prot primary accession no. P08118) as input. This method does not rely on sequence homology. We have subsequently refined the structure using a modification of our method based on an extended set of sequence parameters for the prediction of long-range distance intervals also. Validation of the refined structure for proper stereochemistry and favorable packing environment shows good results. Moreover, functional studies on this structure with respect to its immunogenic and prolactin-suppressing activities are corroborated by the experimental results. [5, 6] Our findings, in turn, offer new and significant structural insights into HSPI's binding activities and related biological and immunogenic functions.

Methodology

We have developed a nonparametric approach to regress the 3D distances between residues (centroids or C_α atoms) as a function of the primary distances and some important features of the primary sequence. [7]

In our model, the unfolded polypeptide chain is represented as a linear sequence of amino acid residues, and each residue is depicted by its van der Waals' sphere.

R.R. Joshi (✉)
Biomedical Engineering Group, Indian Institute of Technology Bombay, Powai, Mumbai: 400 076, India
e-mail: rrj@math.iitb.ac.in
Tel.: +22-576-7485, Fax: +22-572-3480

R.R. Joshi · S. Jyothi
Department of Mathematics,
Indian Institute of Technology Bombay,
Powai, Mumbai: 400 076, India

The 3D distance between the C_α atoms of residue *i* and *j* denoted by d_{ij} is then estimated as a function of the corresponding primary distance p_{ij}, where

$$p_{ij} = r_i + 2r_{i+1} + 2r_{i+2} + ... + 2r_{j-1} + r_j$$

Here r_i denotes the van der Waals' radius of the *i*th residue.

Computational experiments revealed that the primary distances alone are not sufficient to explain the variation observed in the 3D distances in native proteins. Hence, in our model, we consider a few other physical, chemical and geometrical properties of the sequence. In particular, the parameters associated with the size of the sequence, the hydrophobicity of the residues, the four stable clusters of amino acids [8] and certain heuristics on primary–3D distance correlations were used. These parameters were found to be significant with more than 90% confidence.

Since the idea was to develop a prediction method that did not rely on homology, the proteins in the training sample for model estimation were selected randomly from the class of proteins in the PDB up to a size of 150 residues. Identical sequences were discarded. The resulting training sample had 93 proteins. The proteins in this training set are such that much less than 1% of the pairwise sequence alignments show more than 40% identity.

The short and medium-range distances $\{d_{ij} | j–i| \leq 4\}$ were then estimated as smooth additive functions [9] of the above sequence parameters. Also, to lend compactness to the structures, certain long-range distance restraints, obtained by imposing compactness and hydrophobic core building heuristics using the theoretical results on the radius of gyration and hydrophobic residue probability distribution, [10] were also estimated. The distance estimates from the nonparametric regression model along with these heuristic estimates of long-range distance intervals, when used in a Distance Geometry program **dgsol**, [11] were found to give rise to compact native-like structures.

The performance of our method was validated and was found to be better than other extensively used distance-based computational approaches. A comparative study of the protein structures predicted by our method with that determined by a few other distance-based protein structure computation methods like DRAGON and X-PLOR [12] showed that our method performs better in terms of the quality and accuracy of the structures and also in terms of its computational efficiency. [7] Compared with the high resolution X-ray crystallographic and NMR structures, the global RMSD was found to lie between 5 Å and 9 Å for proteins of sizes ranging from 70 to 150 residues. It was also seen that adding just a very small number of long-range contacts (in the form of simulated NMR restraints) vastly improves the accuracy of the calculated structures. Moreover, the local geometry predicted by our method was also found to be stable and consistently accurate when compared to the results of some of the threading methods. [13]

Modified algorithm

We have extended and refined this approach so that short-, medium- and long-range effects of the sequence are extracted in an optimal way for the estimation of short-, medium- and long-range inter-residue C_α distance intervals. The division of the sequence is again as per the sliding window model. [7] The short- and medium-range distance $\{d_{ij} | j–i| \leq 4\}$ correlations are now discretized into three classes based on the heuristics on secondary structure. The long-range distances $\{d_{ij} | j–i| \geq 20\}$ are discretized into two classes based on the distance being less than or greater than the average distance.

Nonparametric discriminant analysis [14] with a normal kernel on a set of sequence parameters that include local and global measures of hydrophobicity, cluster identity [8] and secondary structure propensity [15] gave the best cross-validated results for the estimation of the class membership in the random training sample of 93 proteins. In the case of the short- and medium-range distances, these estimated correlations are used in the nonparametric regression model for the estimation of the corresponding distance intervals.

The heuristic globular constraints are also refined using β-turn propensity [15] and hydrophilicity profile plots. In particular, the residues in the sequence having β-turn propensity greater than a specified threshold or falling in the regions of local maxima of the hydrophilicity profile plots are constrained to lie on the surface of the protein.

The C_α trace of the protein is obtained by incorporating these distance interval estimates in the distance geometry algorithm, **dgsol**. [11] The distinct structures obtained from **dgsol** are further refined using a probabilistic distance geometry program using the posterior probabilities from discriminant analysis. The optimal solutions are selected based on minimal constraint violations. With this approach, the RMSDs of the structures were found to decrease significantly – the RMSDs in the validation samples now vary from 4 to 7 Å. The mathematical and computational details of this methodology and the complete validation results are being reported elsewhere. [16]

Structure prediction of HSPI

We had earlier predicted the structure of HSPI using only the short-range distance estimates along with the globular constraints. [3] Certain disulphide constraints between cysteine pairs were tried out using crude estimates of probabilities [17, 18] on such bonds. Three different sets of disulphide connectivities were tried. Out of the three optimal solutions thus obtained, the one that showed uniqueness and stability with respect to the computation of atomic coordinates with MaxSprout [19] and the presence of hydrogen bonds was selected as the most probable structure of HSPI. The variation in the RMSD of this structure with the other members of this set was between 3.9 and 5.6 Å.

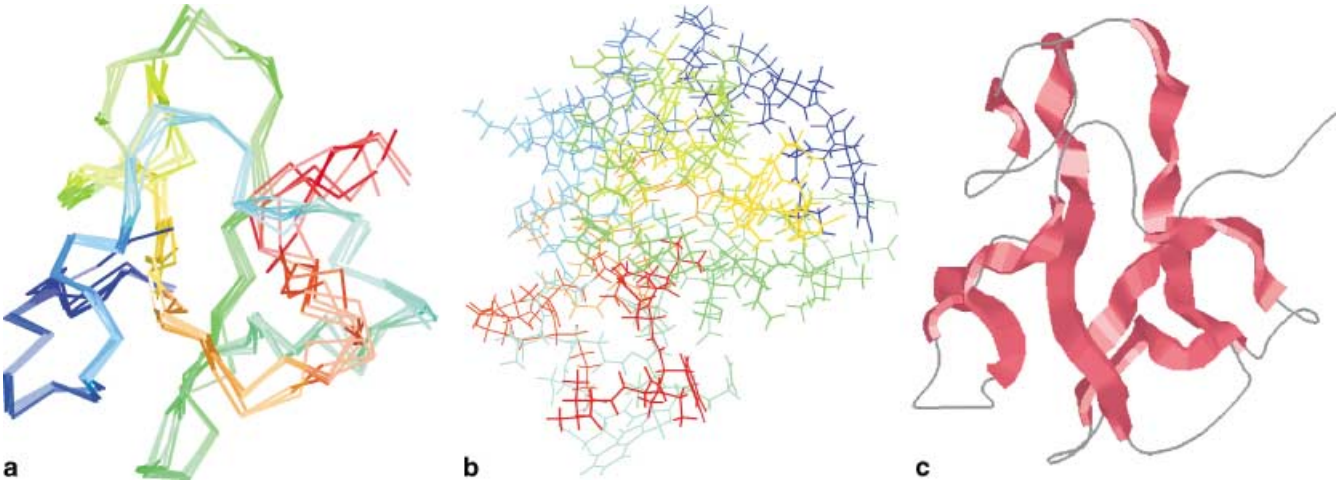

Fig. 1 **a** The superposition of the C_α structures on the overall best structures of HSPI. Figures are color coded in a smooth range from *blue* (N-terminus) to *red* (C-terminus). All molecular structures have been drawn using RASMOL, a molecular graphics visualization package available at http://www.bernstein-plus-sons.com/software/RasMol_2.7.1. **b** Full Atomic Model for the energy-minimized structure of HSPI colored in a smooth range from the N-terminus (*blue*) to the C-terminus (*red*). **c** The energy-minimized structure for HSPI showing the β-strands in *pink* ribbons

This structure did agree with the experimental observation that HSPI may be a predominantly β-sheet structure [1] and that the two tryptophan residues (TRP32 and TRP92) could be highly solvent-exposed. [20] This structure, however, did show a small α-helix (residues 69–72). This was contradictory to the available NMR results. [1] Also the number of hydrogen bonds was 16, which was on the lower side for proteins of this length when compared to high-resolution structures. [19] Further, the stereochemical and nonbonded interaction tests did not confirm it to be a well-refined structure.

The above observations motivated us to look again into the structure of HSPI. In particular we wanted a more accurate estimate of the disulphide constraints.

Refined structure of HSPI

Interestingly, the structure obtained by us now is a significant refinement of the previous structure. Here we have used the modification of the 3D distance-estimation procedure as described in the previous section.

366 short- and medium-range distance constraints estimated by the nonparametric regression model and 122 long-range distance constraints from the discriminant analysis and globular heuristics were obtained. From 300 runs of **dgsol** only nine distinct solutions were obtained. The pairwise RMSD between these solutions is also always less than 1.0 Å.

Noting that HSPI is a protein rich in cysteine residues (10/94) we expect the final 3D structure to be stabilized through the formation of disulfide bridges. The C_α structures now obtained have a very low pairwise variance. The distances between the cysteine residues are also unique. Hence, from the set of distinct solutions the pairwise average distances between the cysteines were found. Cysteine(i) was then paired with cysteine(j) for which

$$\{d(i,j)<d(i,k)\forall k\neq i,j;\{i,j,k\}\in\{2,18,37,40,42,49,50,64,73,87\}\}$$

and $d(i,j)$ denotes the average distance between cysteine(i) and cysteine(j) computed from the set of distinct solutions.

The following cysteine residues were paired based on the above heuristic: (CYS2,CYS50)**5.50**(0.44); (CYS18, CYS64)**10.24**(0.34); (CYS37,CYS87)**9.43**(0.48); (CYS40, CYS42) {**6.27**}(0.003); (CYS49,CYS73) {**16.35**}(0.12). The numbers in boldface represent the average C_α distance and the numbers in parentheses represent the standard deviation of this distance in the set of distinct solutions.

The virtual C_α–disulphide bond length was introduced as an additional constraint and the structures were recomputed using **dgsol**.

Fourteen distinct structures were obtained. These C_α structures were then subjected to the probabilistic refinement protocol. The maximum of the pairwise RMSD is 1.5 Å. These structures are superimposed in Fig. 1a.

From the set of distinct structures obtained, the best in terms of bump distance, globular diameter restraint satisfaction, the theoretical radius of gyration and the residue distribution was chosen. This structure was also found to violate only 2.4% of the total number of 4,654 restraints. Of these, most were the ones with low probability.

The full atomic coordinates (backbone and side chain) for this structure were obtained using the program Max-Sprout. [19] The resulting structure was further refined for proper stereochemistry, especially at the joints, by energy minimization using the AMBER server (http://narfi.compchem.ucsf.edu/). This energy-minimized structure is shown in Fig. 1b. This structure was used for all further studies. Figure 1c also shows this structure with the arrangement of the β-strands in pink ribbons.

Results

Validation of the predicted structure of HSPI

The structure for HSPI obtained by the above method is again a compact, predominantly β-sheet structure with 52% of the residues occupying a β-strand conformation and with no residues in the α-helical conformation. Significantly, the secondary structure predicted by our method is in close agreement with the experimental observation that HSPI may contain 50 to 70 residues in the anti-parallel β-sheet conformation and no residues in the α-helix. [1] We also note that some of the other computational secondary structure prediction algorithms [21] predict only 12 to 17 residues to be in the β state (see Table 1). But the average reliability (as given in the output of some of these Internet programs) of a β-prediction by these methods is just 60% for this protein.

The number of hydrogen bonds in the structure predicted by us is 33, which indicates a stable structure with respect to the threshold of 27 for a protein of this size.

The radius of gyration for this structure is 11.24 Å and 54 residues are part of its hydrophobic core. It may be noted that the theoretical radius of gyration for HSPI is 12.37 Å. Also theoretically, it is expected to contain about 52 residues in its hydrophobic core. [10]

The solvent accessibilities of the residues were computed using a solvent accessibility program written by Gerstein. [22] See Fig. 2 for the solvent accessibility plot. Here, we again note that the two tryptophan residues – TRP32 and TRP92 – are solvent-exposed (>30% accessible surface area) in accord with the experimental results of Krishna et al. [20]

We have tested and validated the structure for good stereochemistry and packing using the WHAT_IF and PROCHECK suite of programs. [23] The overall *G* factor, which is a carefully weighted average of all the tests performed by PROCHECK, is –0.83, which indicates an overall refined structure. The major contribution that makes the *G* factor slightly less than the optimal value of –0.50 comes from the deviation from planarity for the peptide bonds (standard deviation 17.4 Å). All the other tests (bond lengths, bond angles, improper torsion angles and χ_1 χ_2 correlations) show good *Z* scores.

The predicted structure was further evaluated for the pattern of non-bonded interactions using the ERRAT [24] program. This gives a plot of the value of the error function versus the position of a nine-residue sliding window. This plot also provides confidence limits based on comparison with statistics from highly refined crystal structures. Around 90% of the residues here are found to have favorable non-bonded contacts (gray and white bars). See Fig. 3.

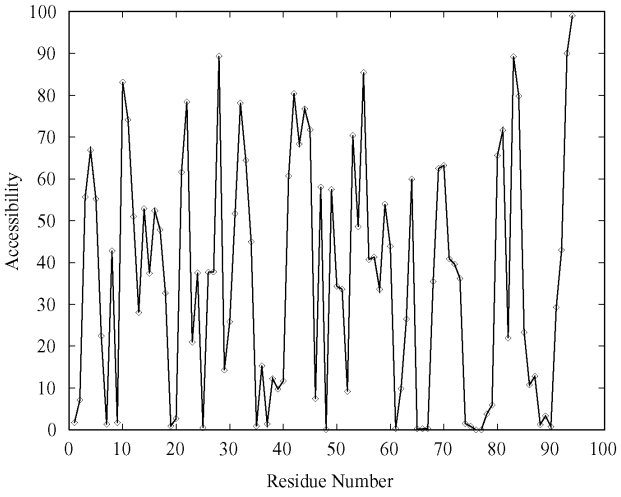

Fig. 2 The percentage solvent accessibility plot for HSPI. Residues with more than 30% accessible surface area are termed as accessible residues

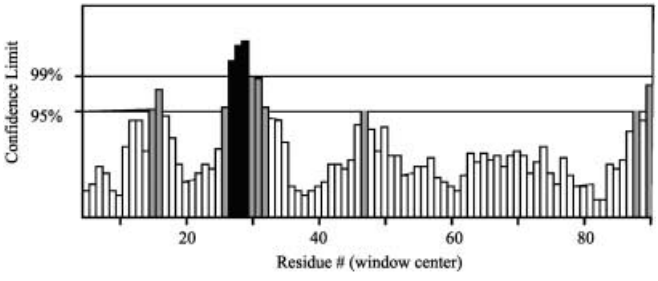

Fig. 3 The output of the ERRAT verification – regions of the structure that can be rejected at the 99% confidence level are *black* and the regions that can be rejected at the 95% confidence level are *gray*

Functions of HSPI

Immunogenic properties

Experimental studies have demonstrated that HSPI, which has FSH suppressing activity, is a sperm-coating antigen with immunocontraceptive potential. [25] Antibodies to HSPI have been found to cause sperm agglutination and impairment of cervical mucus penetration and sperm–egg attachment, severely affecting fertility. A set of polyclonal antibodies against intact inhibin and its se-

Table 1 A comparison of the number of residues in HSPI occupying the different secondary structural states for our method and other secondary structure prediction methods

Number of residues in	Our method	jnet	nnssp	Phd	pred	jpred
α-Helix	0	0	0	0	0	0
β-Extended	49	17	10	15	10	12
Turns and coils	45	77	84	79	84	82

Table 2 The results of docking the different epitopes/intact HSPI on the CDR of IgG. Binding energies are as output by GRAMM

Ligand	Binding sites on the ligand	Binding sites on the receptor	Binding Energy
CE1: {3–5, 41–60, 68–73}	{5, 52, 54, 59}	{45L, 27H, 47H, 97H}	–123
CE2: {8–18, 21–28, 41–60}	{8, 11, 25, 46, 47, 50, 51, 58}	{34L, 45L, 99L, 100L, 29H, 47H, 48H, 51H, 58H, 97H}	–128
CE3: {21–28, 41–60, 92–94}	{21, 23, 24, 25, 26, 27, 46, 47}	{29L, 30L, 31L, 49L, 50L, 27H, 98H, 100H, 101H, 102H, 104H}	–116
CE4: {31–34, 41–60, 92–94}	{31, 34, 41, 46, 48, 52}	{99L, 27H, 29H, 47H, 97H, 98H}	–119
CE5: {41–60, 68–73, 92–94}	{48, 58, 59, 60, 68, 71}	{48L, 49L, 98H, 102H, 104H}	–127
Intact HSPI	{15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 59, 64, 78}	{31L, 45L, 48L, 49L, 31H, 32H, 52H, 53H, 99H, 100H, 101H, 102H, 103H, 104H}	–219

quence-specific fragments – R-17 (residues 1–17) and R-9 (residues 86–94) – were found to affect sperm function by damaging the membrane integrity of the spermatozoa. Antiserum to intact inhibin was found to cause the maximum damage and antiserum to R-9 caused the minimum damage. [5] In further experiments, it was observed from the acrophilicity and hydrophobicity profiles of the protein that the N-terminal 1–17 (R-17) and C-terminal 67–94 (R-28) peptides mimic the immunobiological activity of HSPI. [6]

In order to test our structure against these findings and to shed more light on the immunogenic properties of HSPI, conformational epitopes of HSPI were identified using the algorithm of Kolaskar and Kulkarni-Kale. [26] This algorithm is based on the solvent accessibility and antigenic propensity of the residues.

These conformational epitopes as well as the entire intact inhibin were tested for specific binding to the human IgG. IgG was chosen because the experimental studies were conducted with this immunoglobulin. Each of these conformational epitopes as well as intact HSPI were docked to the CDR of an IgG (PDB Code 2FB4) using the docking program GRAMM. [27, 28] Only the CDR and not the entire IgG molecule was selected as a receptor for docking because the docking algorithms are not reported to be optimal for antigen–antibody docking if the entire Fab of an immunoglobulin is used. Also low resolution docking was preferred to give greater flexibility with respect to the limitation of the docking algorithms. [28] The predicted binding correspondence is shown in Table 2.

It is significant to note that the paratopes (binding sites on the CDR) predicted here are indeed included in the paratopes predicted by a Hopfield network-based program [29] that uses the entire Fab of the IgG as the receptor and that is found to give 84% to 100% correct predictions of paratopes against a conformational/sequential epitope.

The complex of intact HSPI with the IgG is shown in Fig. 4. It is seen that many residues from the R-17 and R-28 segments form part of the conformational epitopes of HSPI, in concordance with the experimental observations. From the large difference in the binding energy between the docking of the epitopes and the docking of intact HSPI, one may also conclude that the other epitope sites (apart from R-17 and R-28) predicted here would contribute to enhancing the immunocontraceptive potential of HSPI.

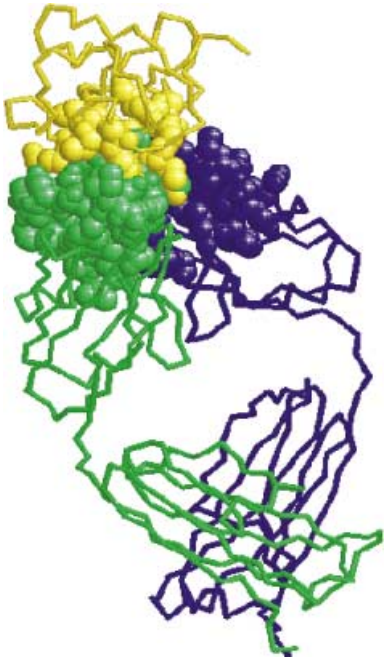

Fig. 4 HSPI–IgG complex. The light and heavy chains of IgG are colored *blue* and *green* respectively. HSPI is colored *yellow*. The CDR region and the binding sites in HSPI are depicted in space-fill representation. Only the backbone is shown for the other residues

Prolactin suppression

Lactation is known to be caused by the prolactin hormone binding to its receptor. [30] The presence of prolactin has been observed in the seminal plasma. [31] Though HSPI has been found to suppress lactation by acting on the prolactin hormone, [32] the mode of action is still unknown. We have conducted docking experiments of HSPI with prolactin hormone (data from PDB file 1AN3) using the GRAMM program. The results

Table 3 The residues in contact in the HSPI–prolactin and the Prolactin–prolactin receptor complexes. The residues in prolactin, which are common in both the complexes, are underlined

Ligand (sites)	Receptor (sites)	Binding energy
HSPI	Prolactin	–406
{8, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 57, 58, 59, 63, 64, 65, 66, 68, 69, 80, 82, 83, 84}	{16, 19, 20, 23, 24, 40, 54, 55, 56, 58, 162, 165, 166, 169, 170, 171, 172, 173, 174, 176, 177, 186, 187, 188}	
Prolactin	Prolactin receptor	–647
{1, 2, 6, 7, 10, 11, 13, 14, 16, 17, 19, 23, 26, 37, 40, 41, 55, 56, 57, 58, 61, 62, 102, 110, 114, 117, 121, 162, 165, 166, 169, 170, 173, 174, 176, 177, 180, 187, 188}	{19B, 20B, 44B, 68B, 71B, 72B, 73B, 74B, 75B, 76B, 78B, 94B, 95B, 96B, 97B, 98B, 100B, 101B, 103B, 136B, 137B, 138B, 139B, 140B, 41B, 142B, 143B, 145B, 189B, 190B, 19C, 20C, 45C, 69C, 71C, 72C, 73C, 74C, 75C, 76C, 98C, 99C, 100C, 103C, 136C, 138C, 139C, 140C, 141C, 143C, 190C}	

throw valuable light on the route of the action of HSPI in suppressing lactation.

The residues in prolactin in contact with HSPI are shown in Table 3. The table also shows the binding sites of the prolactin–prolactin receptor complex.

The prolactin–HSPI complex is shown along with prolactin–prolactin receptor complex in Fig. 5a and b.

It is interesting to note that there are more than 70% residues in common in these two complexes. Significantly, these also happen to be sites in prolactin that were identified to have maximum binding affinity by mutagenesis studies. [30] These experimental studies have also shown that the mutation of these sites (underlined in Table 3) affects the bioactivity of prolactin. Moreover, the remaining sites on prolactin that bind to the prolactin receptor are low affinity ones and are not liable to cause the formation of the prolactin–prolactin receptor complex if the high affinity ones are blocked.

Our results thus provide a new insight explaining how HSPI suppresses lactation by directly acting on the crucial binding sites of prolactin and preventing its binding to the prolactin receptor. Also, interestingly, all the immunogenic sites of HSPI (Table 2 – binding sites of intact HSPI to CDR) constitute part of the binding sites that suppress the formation of prolactin–prolactin receptor complex. This result further provides a conclusive support to the possibilities of increasing milk production by immunosuppression of HSPI.

Follicle stimulating hormone (FSH) and HSPI – mimics of active core

The in vivo studies with respect to the suppression of FSH by HSPI have so far shown the following –

a. HSPI binds to the pituitary cell membrane but the binding is not affected by LHRH (leutenizing hormone release hormone) or BSA (bovine serum albumin) concentration. [33] However, to the best of our knowledge, no further experimental evidence in this regard is reported.
b. Higher concentration of HSPI and its C-terminal 28 residue (67–94) cyclic peptide suppresses the release of FSH but has no effect on the release of LH (leutenizing hormone). These observations were independent of the presence of LHRH. [34, 35, 36]
c. Immunosuppression of HSPI results in higher concentration of FSH that amounts to spermatogenic arrest. [25] Here, experiments with the epitope R-17 have shown 75% reduction in fertility. [6]

However, the in vitro results in the above regard still remain inconclusive. [37, 38, 39]

The observations (b) and (c) reported successively in several experimental studies lead to these interesting questions: (i) Whether there is feasibility of binding between HSPI and FSH above a certain concentration threshold? (ii) Do similarities exist in the active sites of HSPI and FSH, which may lead to the suppression of FSH levels without influencing the release of FSH via LHRH?

As there is no experimental evidence on the binding of FSH and HSPI, the first possibility seems to be unlikely. However, our structure provides interesting clues with respect to the second possibility.

Binding pocket. It has been suggested that the C-terminal 28-residue R-28 (67–94) cyclic peptide of HSPI may have the active core of the protein. [34] We used the cavity detection algorithm of Stahl et al. [40] to find the binding pockets in HSPI. It is indeed interesting to note that all the residues from the region 67–94 are residues forming the lining (border) to the deep surface cavity and hence capable of participating in binding activities. (See Fig. 6a.)

Structural and functional similarity of the binding pocket. We have identified a set of 25 consecutive residues (residues 29–53, chain B) in FSH (PDB code 1FL7), which have close structural similarity with the cyclic 70–94 region of HSPI (see Fig. 6b). Interestingly, the fold of R-17 (residues 1–17) in HSPI did not have any structural similarity with any portion of FSH.

Thus, it appears that the active core R-28 might have a different effect on FSH suppression activity than R-17. We

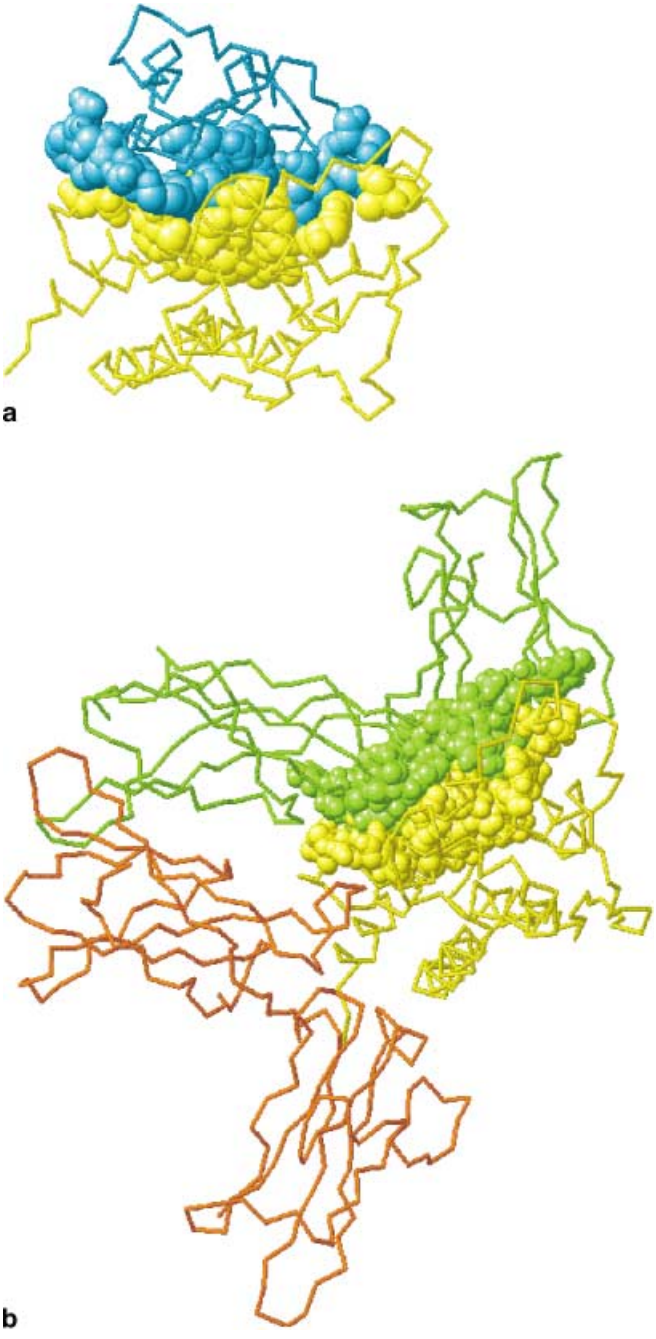

Fig. 5 **a** Prolactin–HSPI complex. Prolactin is colored *yellow* and HSPI in *cyan*. **b** Prolactin–prolactin receptor complex. Prolactin is colored *yellow* and the two chains of the receptor are colored *red* and *green*. The common binding residues in prolactin and the corresponding residues they bind with in HSPI and the prolactin receptor are depicted in space-fill representation. For the sake of clarity, only the backbone is shown for the other residues.

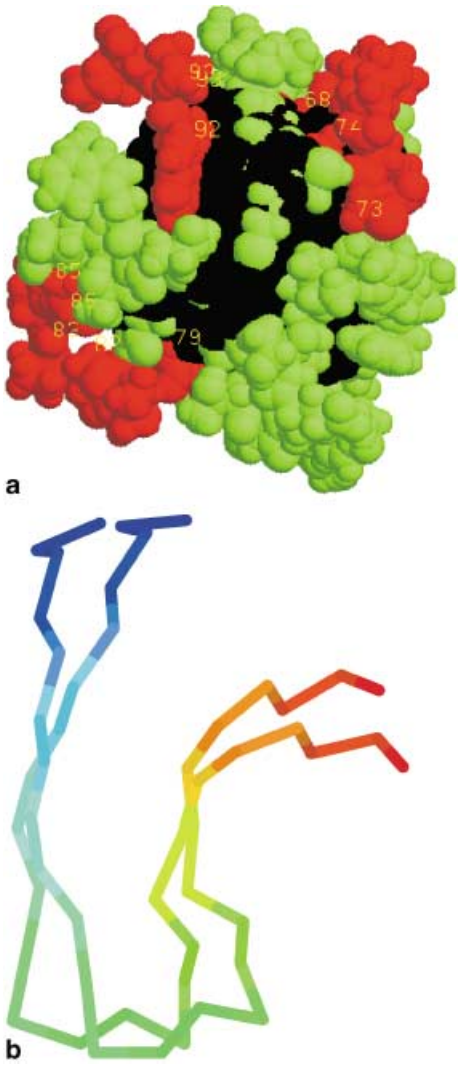

Fig. 6 **a** The cavities in HSPI. The region colored *black* forms the cavity. The residues from 67 to 94 are marked in *red*. These are also seen to form part of the lining of the cavity. A few of these residue numbers are labeled in *yellow*. **b** The superposition of the cyclic fragment 70–94 of HSPI with the fragment 29–53 in the chain B of FSH. The color coding is *blue* (N-terminus) to *red* (C-terminus). The RMSD between the two structures is 4.10 Å. However, if only the portions 70–90 of HSPI and 29–49 of FSH are considered the RMSD reduces to 2.87 Å

have also found that the structurally similar pockets 70–94 in HSPI and 29–53 in FSH share similar charge, polarity and hydrophilicity profiles – especially for a three-residue moving average window, although there is no sequence similarity between the two. This structural and (biophysical) functional similarity is most significant in the portion 76–94 of HSPI and the corresponding 35–53 segment of FSH. This observation favors the possibility that structure function mimicking of the active core of HSPI with a surface-exposed active site of FSH might trigger a false recognition for the FSH secreting system and thus lead to a suppression in FSH level in the presence of free HSPI.

Discussion

Ab initio structure determination of biologically important proteins and identification of their active sites and functional properties offers promising applications in genomics–proteomics research. This assumes greater relevance and significance in the case of proteins like HSPI

that lack sequence homology with known proteins and that are difficult to be studied by NMR and crystallography. Our method, which is based on nonparametric regression and knowledge-based heuristics without relying on sequence homology, contributes in this direction.

We have predicted the structure of HSPI using this approach. The testing of the predicted structure through structure validation programs shows good stereochemistry and packing interactions. Interestingly, as for its sequence, this structure too does not have any homology (as shown by DALI [41]) with the 3D structures available in PDB at present. This structure also shows interesting properties in concordance with the experimental observations – including its antigenic sites and binding with IgG.

Our results on prolactin–HSPI binding indicate the role of HSPI in the suppression of lactation. Identification of common binding sites on the bioactive regions of prolactin provides new insight explaining how HSPI suppresses lactation by directly acting on the crucial binding sites of prolactin and preventing its binding to the prolactin receptor. This gives new directions for research into the possibility of enhancing milk production and eliminating the abnormalities triggered by inhibition of lactation through immunization of inhibin.

The experiments on the immunocontraceptive role of HSPI so far have been reported on the peptide R-17 only. Experiments on rats have shown 75% reduction in fertility. [6] Our results also show the importance of other epitope sites in HSPI–IgG (CDR) binding. In particular, it would be interesting to experiment with the modified peptides – for example the one that consists of the residue numbers 8–28 (see Table 2).

Despite extensive experimental research, the nonlinear feedback interaction between FSH and HSPI still remains a puzzle. Our results offer interesting clues with regard to the active core on HSPI consisting of the cyclic peptide R-28. Interestingly, it is this peptide, and not its linear counterpart, which produced FSH suppression in vivo. [36]

As our results show, this cyclic group does not seem to have any immunogenic active role (see Table 2). To the best of our knowledge, no experimental result is yet reported on the effect of immunosuppression of this region of HSPI, whereas experimental research does show that immunosuppression of the R-17 region enhances the release of FSH.

It would thus be important to experiment with both R-17 (residues 1–17) and cyclic R-28 (residues 67–94) to see their combined effect on the reduction of FSH release. The structure–function mimicking of the active core might counter the impact of immunosuppression of R-17 on FSH release. It is interesting to note that, as reported in their recent paper, [6] Vanage et al. are proceeding to experiment on such combined effects.

Acknowledgements Part of this research has been funded by the Dept. of Biotechnology (DBT), Govt. of India through a project undertaken by Prof. Rajani R. Joshi. The author thanks the DBT for the same. The co-author thanks the Council of Scientific and Industrial Research, Govt. of India for the financial support (Senior Research Fellowship) in undertaking this research work.

References

1. Chary KVR. Rastogi VK, Govil G, Seema VG, Sheth AR (1993) Curr Sci 65:786–788
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
3. Jyothi S, Joshi RR (2000) Protein Pept Lett 7:167–174
4. Rastogi VK, Chary KVR, Govil G (1997) Curr Sci 72:69–71
5. Mehta MK, Sheth AR (1994) Arch Androl 33:129–136
6. Vanage GR, Mehta PB, Moodbidri SB, Iyer KS (2000) Arch Androl 44:11–21
7. Jyothi S, Joshi RR (2001) Comput Chem 25:283–299
8. Stanfel LE (1996) J Theor Biol 183:195–205. DOI 10.1006/jtbi.1996.0213
9. Buja A, Hastie T, Tibshirani R (1989) Ann Stat 17:453–555
10. Andrzej K, Skolnick J (1997) J Chem Phys 107:953–964
11. More JJ, Zhijun Wu (1999) J Global Optim 15:219
12. Aszodi A, Gradwell MJ, Taylor WR (1995) J Mol Biol 251:308–326. DOI 10.1006/jmbi.1995.0436
13. Sippl M (1990) J Mol Biol 213:859–883
14. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, New York
15. Chou PY, Fasman GD (1974) Biochemistry 13:211–222
16. Jyothi S, Joshi RR (2001) in preparation
17. Thornton JM (1981) J Mol Biol 151:261–287
18. Harrison PM, Sternberg MJE (1996) J Mol Biol 264:603–623. DOI 10.1006/jmbi.1996.0664
19. Holm L, Sander C (1991) J Mol Biol 218:183–194
20. Krishna MMG, Rastogi VK, Periasamy N, Chary KVR (1997) J Phys Chem B 102:5520–5528
21. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) Bioinformatics 14:892–893
22. Gerstein M (1992) Acta Crystallogr, Sect A 48:271–276
23. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G (1998) CABIOS 14:523–528
24. Colovos C, Yeates TO (1993) Protein Sci 2:1511–1519
25. Sheth AR, Moodbidri SB, Garde SV, Vanage GR (1992) Indian J Exp Biol 30:1012–1016
26. Kolaskar AS, Kulkarni-Kale U (1999) Virology 261:31–42. DOI 10/1006/viro.1999.9858
27. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992) Proc Natl Acad Sci USA 89:2195–2199
28. Vakser IA (1995) Protein Eng 8:371–377
29. Joshi RR (2001) Protein Pept Lett 8:257–265
30. Halaby D, Thoreau E, Djiane J, Mornon JP (1997) Proteins Struct Funct Genet 27:459–468
31. Dattatreyamurthy M, Sheth AR (1977) Mol Cell Endocrinol 3:253–259
32. Moodbidri SB, Hurkadli KS, Sheth AR (1989) Indian J Exp Biol 27:214–216
33. Vanage GR, Sheth AR (1982) Indian J Exp Biol 20:45–47
34. Arbatti NJ, Seidah NG, Rochemont J, Escher E, Sheth AR, Chretien M (1985) FEBS Lett 181:57–63
35. Mundle SD, Sheth NA (1991) Indian J Exp Biol 29:310–314
36. Mahale SD, Sheth AR, Iyer KS (1993) Int J Pept Protein Res 42:132–137
37. Kohan S, Froysa B, Cederlund E, Fairwell T, Lerne R, Johansson J, Khan S, Ritzen M, Jornvall H, Cekan S (1986) FEBS Lett 199:242–248
38. Iyer KS, Mahale SD, Hurkadli KS, Sheth AR (1989) Indian J Exp Biol 27:10–13
39. Hurkadli KS, Mahale SD, Iyer KS, Sheth AR (1989) Life Sci 45:1357–1363
40. Stahl M, Taroni C, Scheider G (2000) Protein Eng 13:83–88
41. http://www2.ebi.ac.uk/dali